

SEARCH REQUEST FORM

4/26/82
120

Requestor's
Name

Rita Mitra

Serial

Number

09-556246

Date

4-25-82

Phone

608-1211

Art Unit

1653

Rm 9803

MB 9801

Search Topic

Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevant citations, authors, keywords, etc., if known. For sequences, please attach a copy of the sequence, or you may include a copy of the broadest and/or most relevant claim(s).

Please search

AA1

Seg 1, aa 1-24 plus aa 200-1404 (delete aa 25-199)

AA2

Seg 1, aa 1-25 plus 67-1404 (delete aa 26-66)

AA3

Seg 1, aa 1-25 plus aa 67-106 plus aa 200-1404
(delete aa 26-66 and 107-199)

aa 200-1140 of Seg 1

Seg 1 aa 25-end (delete aa 1-24)

AA4

Seg 1 aa 1-66 plus 105-end (delete aa 67-104)

AA5

Seg 1, aa 1-156 plus aa 200-1404 (delete aa 157-199)

AA6

Seg 1, aa 1-106 plus aa 200-1404 (delete aa 107-199)

AA7

Seg 1, aa 200-1167

AA8

Seg 1, aa 200-1212

AA9

Seg 1, aa 200-1263

Point of Contact

Toby Port

Technical Info Specialist

CM16A04

703-308-3634

STAFF USE ONLY

Date completed

4/29

Searcher

T. Port + B. Brynner

Terminal time

120

Elapsed time

20

CPU time

Total time

Number of Searches

Number of Databases

Search Site

SMC

CM-1

Pre-S

Type of Search

N.A. Sequence

A.A. Sequence

Structure

Bibliographic

Vendors

IG Suite

STN

Dialog

APS

Geninfo

SDC

DARC/Questel

Other

T6C+26

BEST AVAILABLE COPY

REPEAT	229	235	1.
FT REPEAT	236	242	2.
FT DOMAIN	205	620	CONTAINS THE SER-PRO(4) REPEATS.
FT DOMAIN	499	600	3 X APPROXIMATE TANDEM REPEATS.
SQ SEQUENCE	620 AA;	65406 MW;	641DD2278AB28524 CRC64;

Query Match 7.6%; Score 551; DB 1; Length 620;
Best local Similarity 27.28; Pred. No. 3.6e-15;
Matches 170; Conservative 64; Mismatches 304; Indels 88; Gaps 24;

QY	297	AEATTK-GPALTTP---KEPTTT-----PKFASTTPKE---PTPTTKSAPTTPKEAP	345
DB	24	AEATQYGGYLPVPVTSQPPSSIGLSPSAPTTPPSRGHVSP---RHAPPRHAYPPP	80
QY	346	TTTKSAPTTPKEP-----APTTTKEAPT---TPKEAPTTPKAPATTTKSAPTTPKE	396
DB	81	SHGLPSPVGGPPHURGLPPSRGNPPSPVISFSPHPPSGAGPPSHGCHLPSHQR	140
QY	397	PAPTTPKAPAPT---TPKEAPTTPKEPTTPKAPATTTK---EPAPTTPKEAPTAPKK	451
DB	141	P-----PSPSHGAPPSGGHTPPRGQHPPSHRPSPPSRHGHPPTTYAQPPPTIYS	193
QY	452	PAPTTPKAPAPTTPKAPATTTKSPPTPK---EPAPTTKSAPTTPKAPATTTKS---	506
DB	194	PSPQV---QPPT---YSPPTTHVQPTSPPSRGHQPPTTHRHAPTHRHAPTHQPSPL	249
QY	507	---APTTPKPSPTTKAPATTTPKAPATTTPKAPATTTPKAPATT---TPKEAPTTPK	560
DB	250	RHLPPSPRQOPPTYSPPPPAYAQSPQSPPTYSPPPTYSPPPTSPPIYSPPPAYSPSP	309
QY	561	KAPT---APKEAPTTPKAPTTPKKT---TPTT-----PEKLAPTTPKAPTTPPEL	610
DB	310	PPTPTTSPPPAYSPPTYSPPPTTYLPSSPIYSPPPVYSPPPVYSPPPPTYL	369
QY	611	APTTPKEPTTPKEAPTTPKAAAP---NTPKAPATTTPKAPATTTPKAPTTPKETA	667
DB	370	PPPPSP	427
QY	668	PTTPKAPTTPKAPATTTPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPK	727
DB	428	P-----LPPTYSPPPPAYSPPPPTYSPPPTYSPPPPAYAQPPPTYSPPPPAYSP	481
QY	728	APTTPKAPATTTPKAPTTPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPK	787
DB	482	PPSPPIYSPPPVYSPPPVYSPPPVYSPPPVYSPPPVYSPPPVYSPPPVYSPPPVY	538
QY	788	TAPTTPKAPAPT---TPKAPAPTTPPTTSEVSTPTTKEPTTHKSPDESTPELSAE	845
DB	539	PPRQIHSPPPHWPRTPTTYGQPPSPPTFSAPPRQIHSPPPHWPRTPTTYGQP	598
QY	846	PTPKALENSKPECPVPTTKTAPATKP	871
DB	599	PSP-----PTTYSPPSPPP	612

RESULT 8

NFH_MOUSE	STANDARD;	PRT;	1087 AA.
ID	NFH_MOUSE		
AC	P19246; 061959;		
DT	01-NOV-1990 (Rel. 16, Created)		
DT	01-AUG-1992 (Rel. 23, Last sequence update)		
DT	20-AUG-2001 (Rel. 40, Last annotation update)		
DE	NEUROFILAMENT TRIPLET H PROTEIN (200 KDA NEUROFILAMENT PROTEIN)		
DE	(NEUROFILAMENT HEAVY POLYPEPTIDE) (NF-H).		
GN	NEFH OR NFH.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_taxid=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
FX	MEDLINE=89121513; PubMed=3220257;		

QY	513	EPSPTTTKEAPTTPKAPTTPKKA-----PTTPKAPATTTPKAPATTTPKAPATTTP	567
DB	702	EKPTISTEE---PTTPEETWISTEKSPIMKPTLPTETTTISVEETISTEKLIPM---	757
QY	568	KEAPTTPKAPT-----TPKLTTPTEKLA-----PTTKEKAPATTTPKAPATTTP	617
DB	758	EKPTISTEKTPTTPTTISPEKLTPT---EKLTPKEKPTTPTTPTTISPEKLT---	814
QY	618	PTTPTTPEAPTTPKAAAPNTKAPATTTPKAPATTTPKAPTTPKAPTTPKAPTTP	677
DB	815	KPTISPEKPTISTEK-----PTTPTT-KPTIPT-----ETISTEKLITPT	855
QY	678	TLKEAPTTPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAP	731
DB	856	---EKTPTISPEKLTPTTPTTISPEKLTPTTPTTPTTPTTPTTPTTPTTPTTPTT	911
QY	732	PKEAPTTPKAPTTPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPK	789
DB	912	PPHSPPTA-TGLAALVMSHAPSTPMTSV---ILGTTTTSRSTGMSCP-PNARYESC	966
QY	790	PTTPKAPT	799
DB	967	PASCKSPRS	976

RESULT 7

EXTN_TOBAC	STANDARD;	PRT;	620 AA.
AC	P13983;		
DT	01-JAN-1990 (Rel. 13, Created)		
DT	01-JAN-1990 (Rel. 13, Last sequence update)		
DT	01-MAR-1992 (Rel. 21, Last annotation update)		
DE	EXTENSIN PRECURSOR (CELL WALL HYDROXYPROLINE-RICH GLYCOPROTEIN).		
GN	HRCPT3.		
OS	Nicotiana tabacum (Common tobacco).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;		
OC	Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.		
OX	NCBI_taxid=4097;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=CV_XANTHI; TISSUE=Leaf;		
RX	MEDLINE=90128263; PubMed=2612509;		
RA	Keller B., Lamb C.J.;		
RT	"Specific expression of a novel cell wall hydroxyproline-rich		
RT	glycoprotein gene in lateral root initiation."		
RL	Genes Dev. 3:1639-1646(1989).		
CC	-!- FUNCTION: HAS A SPECIALIZED STRUCTURAL FUNCTION, POSSIBLY IN		
CC	THE MECHANICAL PENETRATION OF THE CORTEX AND EPIDERMIS OF THE		
CC	MAIN ROOT.		
CC	-!- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX.		
CC	-!- PTM: EXTENSINS CONTAIN A CHARACTERISTIC REPEAT OF THE PENTAPEPTIDE		
CC	SER-PRO(4). THE PROLINE RESIDUE IS HYDROXYLATED AND THEN		
CC	GLYCOSYLATED.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
CC	EMBL: X13885; CAA32090.1; --		
DR	PIR: S06733; S06733.		
DR	Repeat; Cell wall; Glycoprotein; Signal; Structural protein;		
KW	Hydroxylation.		
FT	SIGNAL	1	?
FT	CHAIN	70	EXTENSIN.
FT	REPEAT	148	H-A-P-P.
FT	REPEAT	151	H-A-P-P.
FT	DOMAIN	229	2 X 7 AA TANDEM REPEATS OF T-H-R-H-A-P-P.

QY	821	TPTTTTKEPTT-----IHKSPDESTPELSAEPPTPKALENSPKPGVPTTKTTPAAATKPEM	873
DB	701	APDTRPAPGCTAPPAPGVTSA	753
QY	874	TTTAKDKTTERDLRTTPETT-----TAAAPKMKETATITTEKTESKITATTQTQVTSIT	926
DB	754	--PAHGVTSA	807
QY	927	TQDUTTP--FKITLTKTTLAPKVTTKTKITTT--EIMNKPETAKPKDRATNSKATTPK	982
DB	808	PGSTAPPAHGVTSA	867
QY	983	POKPTKAPKKPTSTKKPTMPVRVKPKTTPTPRKMTSTMPELNPTSKIAEAMLQTTTRPN	1042
DB	868	P--GSTAPPAHGVTSA	909
QY	1043	QT--PNSKLVEVNPKESDAGGAGETPHMLLRPHVFMPEVTPDMYLPVNPQGIINPML	1101
DB	910	STAPPAHGVTSA	955
QY	1102	SDETINCNGKPYDGLTLTLRNGTLVAFRGHYFWMLSPPSPPS	1142
DB	956	HNVTSSAGSAGSASTLVHNGTSARATTTTTPASKSTPESIPS	996

RESULT 5

FPI_MYTED

ID

FPI_MYTED

STANDARD;

PRT;

875 AA.

AC

Q25460;

DT

30-MAY-2000 (Rel. 39, Created)

DT

30-MAY-2000 (Rel. 39, Last sequence update)

DT

20-AUG-2001 (Rel. 40, Last annotation update)

DE

ADHESIVE PLAQUE MATRIX PROTEIN (POLYPHENOLIC ADHESIVE PROTEIN) (FOOT PROTEIN 1) (MEPFI) (FRAGMENT).

GN

FPI.

OS

Mytilus edulis (Blue mussel).

OC

Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloidea;

OC

Mytiloidea; Mytilidae; Mytilus.

OX

NCBI_TaxID=6550;

RN

[1]

RN

SEQUENCE FROM N.A.

RX

MEDLINE-91025929; Pubmed-1367451;

RA

Filipula D.R., Lee S.M., Link R.P., Strausberg S.L., Strausberg R.L.;

RT

"Structural and functional repetition in a marine mussel adhesive protein.";

RL

Biotechnol. Prog. 6:171-177(1990).

RN

[2]

RP

PARTIAL SEQUENCE, AND POST-TRANSLATIONAL MODIFICATIONS.

RA

MEDLINE-83135732; PubMed-6298211;

RX

Waite J.H.;

RT

"Evidence for a repeating 3,4-dihydroxyphenylalanine- and hydroxyproline-containing decapeptide in the adhesive protein of the mussel, Mytilus edulis L.";

RL

J. Biol. Chem. 258:2911-2915(1983).

CC

-1- FUNCTION: PROVIDES ADHESIVENESS TO THE MUSSEL'S FOOT. MUSSELS PRODUCE ONE OF THE STRONGEST WATER INSOLUBLE GLUE. THE MUSSEL'S ADHESIVE IS A BUNDLE OF THREADS, CALLED A BYSSUS, FORMED BY A FIBROUS COLLAGENOUS CORE COATED WITH ADHESIVE PROTEINS.

CC

-1- SUBCELLULAR LOCATION: SECRETED.

CC

-1- TISSUE SPECIFICITY: PRODUCED BY THE BYSSAL GLAND.

CC

-1- DOMAIN: ALMOST EXCLUSIVELY COMPOSED OF REPEATS OF A DECAPEPTIDE.

CC

-1- PTM: THE DECAPEPTIDE A-K-P-S-Y-P-P-T-Y-K IS POST-TRANSLATIONALLY MODIFIED AS FOLLOWING: THE SIXTH AND SEVENTH RESIDUES ARE HYDROXYLATED AND THE PENULTIMATE IS A 3,4-DIHYDROXYPHENYLALANINE (DOPA) DERIVED FROM TYROSINE.

CC

CC

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/annotation/send_email_to_license@sib-sib.ch).

"Preoperative diagnosis of thyroid papillary carcinoma by reverse transcriptase polymerase chain reaction of the MUC1 gene.";

Int. J. Cancer 66:55-59(1996).
[11]
SEQUENCE OF 1-89 FROM N.A.
TISSUE=Lung;
MEDLINE=96181716; PubMed=8604237;
RA Yu C.J., Yang P.C., Shew J.Y., Hong T.M., Yang S.C., Lee Y.C.,
Lee L.N., Luh K.T., Wu C.W.;
"Mucin mRNA expression in lung adenocarcinoma cell lines and
tissues.";
Oncology 53:118-126(1996).
[12]

SEQUENCE OF 1-46 FROM N.A.
TISSUE=Breast carcinoma;
RA Buluwela L., Liu Q., Lugmani Y.A., Gomm J.J., Coombes R.C.;
RA Submitted (Oct-1992) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: DIRECT OR INDIRECT INTERACTION WITH ACTIN
CYTOSKELETON.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. A SECRETED FORM
IS ALSO PRODUCED.
CC -!- ALTERNATIVE PRODUCTS: VARIOUS VARIANTS ARE PRODUCED BY ALTERNATIVE
SPLICING.
CC -!- TISSUE SPECIFICITY: ABERRANTLY EXPRESSED IN HUMAN EPITHELIAL
TUMORS, SUCH AS BREAST CANCER.
CC -!- PTM: HIGHLY GLYCOSYLATED (N-AND O-LINKED CARBOHYDRATES AND SIALIC
ACID).
CC -!- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC. IT
VARIES FROM 21 TO 125 IN THE NORTHERN EUROPEAN POPULATION. THE
MOST FREQUENT ALLELES CONTAINS 41 AND 85 REPEATS.
CC -!- SIMILARITY: CONTAINS 1 SEA DOMAIN.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; J05582; AAA60019.1; -
EMBL; M32738; AAA35804.1; -
EMBL; M32739; AAA35806.1; -
EMBL; J05581; AAA59876.1; -
EMBL; M61170; AAB53150.1; -
EMBL; X52229; CAA36478.1; ALT_SEQ.
EMBL; X52228; CAA36477.1; ALT_SEQ.
EMBL; M35093; AAB59612.1; ALT_SEQ.
EMBL; Z17324; CAA78972.1; -
EMBL; Z17325; CAA78973.1; -
EMBL; M31823; AAA35757.1; -
EMBL; S81781; AAD14376.1; ALT_INIT.
EMBL; S81736; AAD14369.1; ALT_INIT.
EMBL; M21868; AAA59874.1; ALT_SEQ.
PIR; B35175; B35175.
PIR; B35175; B35175.
PIR; S10218; S10218.
GlycoSuiteDB; P15941; -
MIM; 158340; -
MIM; 113720; -
InterPro; IPR000082; SEA.
Pfam; PF01390; SEA; 1.
SMART; SM00200; SEA; 1.
PROSITE; PS50024; SEA; 1.
Glycoprotein; Signal; Cytoskeleton; Actin-binding; Transmembrane;
Repeat; Alternative splicing.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 1255 MUCIN 1.
FT DOMAIN 24 1162 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1163 1186 POTENTIAL.
FT DOMAIN 1187 1255 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 181 960 44 X 20 AA TANDEM REPEATS.
FT DOMAIN 1034 1151 SEA.

FT	CARBOHYD	957	957	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	975	975	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1029	1029	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1055	1055	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1133	1133	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	VARSPPLIC	19	19	T -> TATTAPKAT (IN ISOFORM B).	
FT	VARSPPLIC	20	22	MISSING (IN ISOFORM C).	
FT	VARSPPLIC	20	31	MISSING (IN ISOFORM D).	
FT	VARSPPLIC	126	905	MISSING (IN POLYMORPHIC EPITHELIAL ISOFORM).	
FT	VARSPPLIC	1077	1087	FLQYKGGFL -> VSIGLSFPMPL (IN SECRETED ISOFORM).	
FT	VARSPPLIC	1088	1255	MISSING (IN SECRETED ISOFORM).	
FT	CONFLICT	2	2	T -> A (IN REF. 11).	
FT	CONFLICT	134	134	P -> Q (IN REF. 9).	
FT	CONFLICT	154	154	P -> Q (IN REF. 9).	
FT	CONFLICT	1021	1021	S -> T (IN REF. 3).	
FT	CONFLICT	1251	1251	A -> T (IN REF. 3).	
SQ	SEQUENCE	1255	AA; 122072	MW; 5E28DFC4DE7D9A82	CRC64;

Query Match 8.9%; Score 651; DB 1; Length 1255;
Best Local Similarity 27.8%; Pred. No. 9.3e-19;
Matches 295; Conservative 72; Mismatches 460; Indels 234; Gaps 51;

QY	236	TSLVNKETTVETKETTNNKQSTDKKETS	SAKETQSIKTS	SAKD-LAPTSKVLAKPT	294
Db	16	TVLTV-----VTGSGHASSTPGGKET	SATORSVSPS	TEKNAVSTSSVLSHS	65
QY	295	P-KAETTTKGP--ALITPKP-----	-----PPTPKPEASTTPKEPTPTTI	---K	334
Db	66	PGSSSTTQODVTLAPATPAGSGSAATWQ	DVTSVEVTRPALGSTTPPAHDVTSAPDNK	125	
QY	335	SAPTTKPEAPTTKSAPTTKPEAPTTKPE	-----APTTPKEAPTTTKPEAPTTTK	388	
Db	126	PAPGSTAPPAGHVT-SAPDT--RPAGST	APPAGHVT-SAPDT--RPAGSTAPPAGHVT	179	
QY	389	SAPTTKPEAPTTTP-----KKPAPTTP	KPEP-----APTTPKEPTPTTP-----	426	
Db	180	SAPDTRPAGSTAPPAGHVT-SAPDTRPA	PGSTAPPAGHVT-SAPDTRPAGSTAPPAGHVT	239	
QY	427	---KEPAPTTPK-----APTTPKEAP	TAPK-----KPAPTTPKEAPTTPK-----	E	467
Db	240	SAPDTRPAGSTAPPAGHVT-SAPDTRPA	PGSTAPPAGHVT-SAPDTRPAGSTAPPAGHVT	299	
QY	468	PATTTKESPTTPKEAPTT-----TKS	APTTPKEAPTTTKSAPTTPKESPTTPKEP	522	
Db	300	SAPDTRPAGSTAPPAGHVT-SAPDTRPA	PGSTAPPAGHVT-SAPDT--RPAGSTAPP	356	
QY	523	---APTTPKEAPTTTPK-----KPAPT	TPKEAPTTTPK-----EPAPTTPKKAPTAPKEP	570	
Db	357	GVTSAPDTRPAGSTAPPAGHVT-SAPDTR	PAAGSTAPPAGHVT-SAPDTRPAGSTAPP	416	
QY	571	APTTPKETAPTTPKLTPTTPEKLAPTTP	KEPAPTTPPELAPTTPPEPTTPPEEPAPT	630	
Db	417	GVTSAPDTRPAGSTAPPAGHVT-SAPDTR	PAAGSTAPPAGHVT-SAPDT--RPAGSTAPP	472	
QY	631	PKA-----AAPNTPKPEAPTTPK-----	EPAPTTPKEAPTTPKETAPTTPKGTAPTTPKEP	682	
Db	473	PPAGHVT-SAPDTRPAGSTAPPAGHVT	SAPDTRRPAAGSTAP----PAHGVTSAPDTR	528	
QY	683	APTTP-----KKPAPKELAPTTPTTKE	PTSTSDKPAP--TTPKGTAPTTPKEAPT	730	
Db	529	GSTAPPAGHVT-SAPDTRPAGSTAP-	PAHGVTSAPDTRPAGSTAPPAGHVTIS----	APD	583
QY	731	TPKEAPTTTPK-----TAPTTLKEAP	TTTP-----KKPAPKELAPTTPTTKGTSTT	776	
Db	584	TRPAGSTAPPAGHVT-SAPDTRPAGST	APPAGHVT-SAPDTRPAGSTAPPAGHVT	642	
QY	777	SDKPAP--TTPK-----ETAPTTPKE	APTTTP-----KKPAPTTPPETTPPTTSVS	820	
Db	643	DTRPAGSTAPPAGHVT-SAPDTRPAG	STAPPAGHVT-SAPDTRPAGS--TAPPAGHVT	700	

Mon Apr 29 08:35:21 2002

```

CC -1- SUBUNIT: ASSEMBLED INTO MONO-LAYERED CRYSTALLINE ARRAYS.
CC -1- SUBCELLULAR LOCATION: CELL WALL.
CC -1- SIMILARITY: CONTAINS 4 S-LAYER HOMOLGY (SLH) DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X67506; CAA47841.1; --
CC InterPro: IPR001119; SLH.
CC Pfam: PF00395; SLH_3
CC PROSITE: PS01072; SLH_DOMAIN; 2.
CC Cell wall; S-layer; Signal; Repeat.
CC SIGNAL 1 28 POTENTIAL.
CC CHAIN 29 1664 CELL SURFACE GLYCOPROTEIN 1.
CC DOMAIN 36 763 4 X 156 AA APPROXIMATE REPEATS.
CC REPEAT 36 191 1.
CC REPEAT 207 363 2.
CC REPEAT 409 565 3.
CC REPEAT 607 763 4.
CC DOMAIN 771 1377 APPROXIMATE TANDEM REPEATS OF
CC T-P-S-D-E-P.
CC GLY/PRO/SER/THR-RICH.
CC SLH 1 (INCOMPLETE).
CC SLH 2.
CC SLH 3.
CC SLH 4 (INCOMPLETE).
CC DOMAIN 1626 1646
CC SEQUENCE 1664 AA; 178194 MW; 5F396695BA9FE/4B CRC64;
CC
CC Query Match 13.0%; Score 950; DB 1; Length 1664;
CC Best Local Similarity 31.0%; Pred No. 3.8e-30;
CC Matches 312; Conservative 104; Mismatches 361; Indels 230; Gaps 55;
CC
QY 289 VLAKPTTP-KAETTTGKALTTTPKEPTTPPKPEASTTPKEPTTPKSAPTTPKPEAPT 347
DB 758 VVIQAPAKAASDEPIPTDTPSDETPS-----DETPS---DETPSDEPTPSD 804
QY 348 TKSAPTTPKEAPTTPKPEAPTTPKEPA-----PTTTPKEAPTTPKSAPTTPKPEAPTTPK 403
DB 805 EPTSPETPEEPIPTDTPSDETPSDETPSDETPSDETPSDETPSETPEEPIPTDTP 864
QY 404 KPAPTTPKEAPTTPKEPTTPKPEAPTTPKEAPT-TPKEAPTAPKPAKPAKPAK 462
DB 865 SDETPSDETPS--DETPS--DETP-SDETPSETPEEPIPTDTPSDETPSDETP 919
QY 463 TTPKEPA--PTTTPKESPT-TPKEAPTTPKSA-APTTPKEAPT---TTKSAPTTPKES 515
DB 920 SDETPSDETPSDETPSETPEEPIPTDTPSDETPSDETPSDETPSDETPSDETP 979
QY 516 ----PTTTPKEAPT-TPKEAPTTPKPAKPAKPAKPAKPAKPAKPAKPAKPAK 570
DB 980 PSDETPSDETPSETPEEPIPTDTPSDETPSDETPSDETPSDETPSDETPSDETP 1039
QY 571 APT---TPKETAPTTPKLTTPKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAK 623
DB 1040 TPDSETPSE---TPKEEPIPTDTPSDETPSDETPSDETPSDETPSDETPSDETP 1094
QY 624 EEPAPTTPKAAAPNTTPKEAPTTPKEPA--PTTTPKEAPT-TPKETAPTTPKGTAPT 680
DB 1095 EEPITDTPSDETPSDETPSDETPSDETPSDETPSDETPSDETPSETPEEPIPTDTP 1154
QY 681 EPAPTTPKPAKPAKELAPTTTKEPT-STTSOKPAPTTPKGTAPTTPKPAKPAKPAK 739
DB 1155 EPTPS--DETPSD-EPTPSDETPSETPEEPIPTDTPSDETPSDETPSDETPS--DETPS- 1208
QY 740 PKGTAPTTPKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAK 795
DB 1209 ---DETPSDETPSETPEEPI-----PTDTPSDETPSDETPSDETPSDETPSDE 1253

```

```

QY 509 TTPKESPTTTPKEAPTTPKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAK 568
DB 1703 TTPSSBI-TTTPTPSSSTTPSPPTTTPSPPTTTPSPPTTTPSPPTTTPSPPTTTP 1761
QY 569 EPAPTTP-----KETAPTTP----- 583
DB 1762 PPSITDTPSPSTTPPTTPCPVCLNWTGWLDSGKNEHKPGDRELIGDVGCGPAAANI 1821
QY 584 ---KKLTP----- 590
DB 1822 SCRATWYPDVPIGOLGQTVVCDVSVGLICKEDQKPGGVPIMAFCLNVEINVOCCEVTQ 1881
QY 591 PEKLAATTPEKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAK 649
DB 1882 PTTMTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNT 1941
QY 650 P-----APTTPKEAPT-TPKETAPTTPKGTAPTTPKPAKPAKPAKPAKPAKPAK 703
DB 1942 PTTTPTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTT 2001
QY 704 TSSTSDKAP-TPPKGTAPTTPKEAPTTPKEAPTTPKGTAPTTPKPAKPAKPAKPAK 762
DB 2002 TPTPTGTQTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 2060
QY 763 ELAPTTPKGTSTSDKAP-TPKETAPTTPKEAPTTPKPAKPAKPAKPAKPAKPAKPAK 821
DB 2061 PTTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNT 2119
QY 822 PTTTKEPTTHKSDSTPELSAEPKPALENSKPEP-----GVPT-TKTPAATK 870
DB 2120 PTGTQTPTT---TPTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNT 2176
QY 871 PEMTTAKDKTTERDLR---TTP---ETTAAKMT---KETATTTKTESKITATTTQV 922
DB 2177 ITTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNT 2235
QY 923 TSTTTQDTPPKITLTKTTLAPKVVTT-KKTTTTEIMMKPEBETAKPBRANSKATTP 981
DB 2236 TGTQPTTTP---ITT---TTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTT 2291
QY 982 KPQETKAP-KKPTSTKPKMPVRKPKTTPTRKMTSTMPELNP---TSRIAEMLOT 1037
DB 2292 ITTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNT 2351
QY 1038 TTR-ENQTPNSKLVBNPKSEDAGBETHMLLRPHVFMPEVTP 1082
DB 2352 GTQTPPTTPTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNT 2392

```

```

RESULT 2
SLP1_CLOTH STANDARD; PRT; 1664 AA.
AC Q06852;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE CELL SURFACE GLYCOPROTEIN 1 PRECURSOR (OUTER LAYER PROTEIN B) (S-LAYER
DE PROTEIN 1).
GN OLPB.
OS Clostridium thermocellum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OC NCBI_TaxID=1515;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCIB 10682;
RX MEDLINE=9320931; PubMed=8458832;
RA Fujino T., Beguin P., Aubert J.-P.;
RT "Organization of a Clostridium thermocellum gene cluster encoding the
RT cellulosomal scaffolding protein CipA and a protein possibly involved
RT in attachment of the cellulosome to the cell surface.";
RL J. Bacteriol. 175:1891-1899(1993).

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:32:37 ; Search time 62.75 Seconds

(without alignments)
798.154 Million cell updates/sec

Title: AA6

Perfect score: 7294

Sequence: 1 MAWKTLPIYLILLLSVEFIQ.....ARAITRSGOTLSKVWYNCP 1366

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1171	16.1	5179	1 MUC2_HUMAN	Q02817 homo sapien
2	950	13.0	1664	1 SLPI_CLOTH	Q06852 clostridium
3	810.5	11.1	1367	1 AMYH_YEAST	P08640 saccharomyc
4	651	8.9	1255	1 MUC1_HUMAN	P15941 h mucin 1 p
5	559	7.7	875	1 FPL_MYTCD	Q25460 mytilus edu
6	555.5	7.6	2700	1 ZAN_MOUSE	Q9Y493 homo sapien
7	551	7.6	620	1 EXTN_TORAC	P13983 nicotiana t
8	551	7.6	1087	1 NFH_MOUSE	P19246 mus musculus
9	530.5	7.3	1162	1 TCNA_TRYCR	P23253 trypanosoma
10	530	7.3	865	1 CPN_DROME	Q02910 drosophila
11	518.5	7.1	872	1 FPL_MYTGO	Q25434 mytilus cor
12	517.5	7.1	831	1 NFH_RAT	P16884 rattus norv
13	509.5	7.0	662	1 MUC1_XENLA	P13983 nicotiana t
14	503.5	6.9	1020	1 NFH_HUMAN	P19246 mus musculus
15	502	6.8	1970	1 RPB1_MOUSE	P23253 trypanosoma
16	498.5	6.8	1020	1 RPB1_MOUSE	Q02910 drosophila
17	493.5	6.8	467	1 BAT2_HUMAN	Q25434 mytilus cor
18	490	6.7	2142	1 SSP2_PLAYO	P16884 rattus norv
19	488.5	6.7	826	1 EXTN_MAZE	Q05049 xenopus lae
20	475.5	6.5	267	1 ZAN_MOUSE	P12036 homo sapien
21	471.5	6.5	5376	1 ZAN_MOUSE	P08775 mus musculus
22	468.5	6.4	634	1 HWPI_CANAL	P14144 cricetus
23	467	6.4	817	1 VRP1_YEAST	Q01443 plasmodium
24	454	6.2	797	1 VGLX_HSVB	P46593 candida alb
25	454	6.2	1161	1 YJ9P_YEAST	P37370 saccharomyc
26	453.5	6.2	1229	1 N121_HUMAN	P28968 equine herp
27	448.5	6.1	670	1 VGS0_HSV11	P47179 saccharomyc
28	444.5	6.1	2774	1 MAPA_RAT	Q00130 ictaluriid h
29	443.5	6.1	1794	1 YAVI_SCHPO	P34926 rattus norv
30	442.5	6.1	751	1 FPL_MYTGA	Q10172 schizosach
31	439.5	6.0	1083	1 T2D3_HUMAN	Q27409 mytilus gal
32	436.5	6.0	1411	1 TCDF_HUMAN	Q00268 homo sapien
33	435.5	6.0	3256	1 K167_HUMAN	Q13428 homo sapien
					P46013 homo sapien

34	432.5	5.9	439	1 XP2_XENLA	P17437 xenopus lae
35	432	5.9	3164	1 TEGU_HSV11	P10220 herpes simp
36	428	5.9	2517	1 NCR2_HUMAN	Q9Y618 h nuclear r
37	427.5	5.9	1251	1 YQUS_CABEL	Q09550 caenorhabdi
38	426	5.8	2715	1 TRX2_HUMAN	Q9umm6 homo sapien
39	425.5	5.8	1185	1 DRPL_HUMAN	P54239 homo sapien
40	424.5	5.8	1125	1 MAP4_MOUSE	P27546 mus musculu
41	424	5.8	307	1 SGS3_DROME	P02840 drosophila
42	423.5	5.8	3421	1 TEGU_HSVB	P28955 equine herp
43	422.5	5.8	1183	1 DRPL_RAT	P24258 rattus norv
44	419.5	5.8	2476	1 ZAN_FIG	Q28983 sus scrofa
45	411.5	5.6	907	1 VGP3_EBV	P03200 epstein-bar

ALIGNMENTS

RESULT	ID	MUC2_HUMAN	STANDARD:	PRT:	5179 AA.
1	AC	Q02817	Q14878		
	DT	01-JUN-1994	(Rel. 29, Created)		
	DT	01-NOV-1997	(Rel. 35, Last sequence update)		
	DT	20-AUG-2001	(Rel. 40, Last annotation update)		
	DE	MUCIN 2 PRECURSOR (INTESTINAL MUCIN 2).			
	GN	MUC2 OR SMUC.			
	OS	Homo sapiens (Human).			
	OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
	OC	Mammalia; Euthera; Primates; Catarrhini; Homiidae; Homo.			
	OX	NCBI_TaxID=9606;			
	RN	[1]			
	RP	SEQUENCE FROM N.A.			
	RC	TISSUE=Intestine;			
	RX	MEDLINE=94132002; PubMed=8300571;			
	RA	Gum J.R. Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;			
	RT	"Molecular cloning of human intestinal mucin (MUC2) cDNA.			
	RT	Identification of the amino terminus and overall sequence similarity			
	RT	to prepro-von Willebrand factor.";			
	RL	J. Biol. Chem. 269:2440-2446(1994).			
	RN	[2]			
	RP	SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.			
	RC	TISSUE=Colon;			
	RX	MEDLINE=93016075; PubMed=1400449;			
	RA	Gum J.R. Jr., Hicks J.W., Toribara N.W., Rothe E.-M., Lagace R.E.,			
	RT	Kim Y.S.;			
	RT	"The human MUC2 intestinal mucin has cysteine-rich subdomains located			
	RT	both upstream and downstream of its central repetitive region.";			
	RL	J. Biol. Chem. 267:21375-21383(1992).			
	RN	[3]			
	RP	SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.			
	RX	MEDLINE=91358717; PubMed=1885763;			
	RA	Toribara N.W., Gum J.R. Jr., Culhane P.J., Lagace R.E., Hicks J.W.,			
	RT	Petersen G.M., Kim Y.S.;			
	RT	"MUC-2 human small intestinal mucin gene structure. Repeated arrays			
	RT	and polymorphism.";			
	RL	J. Clin. Invest. 88:1005-1013(1991).			
	CC	- FUNCTION: COATS THE EPITHELIA OF THE INTESTINES, AIRWAYS, AND			
	CC	OTHER MUCUS MEMBRANE-CONTAINING ORGANS. THOUGHT TO PROVIDE A			
	CC	PROTECTIVE, LUBRICATING BARRIER AGAINST PARTICLES AND INFECTIOUS			
	CC	AGENTS AT MUCOSAL SURFACES.			
	CC	- SUBUNIT: MULTIMERIC.			
	CC	- SUBCELLULAR LOCATION: COLON, SMALL INTESTINE, COLONIC TUMORS,			
	CC	BRONCHUS, CERVIX AND GALL BLADDER.			
	CC	- PM: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR			
	CC	INTERCHAIN DISULFIDE BONDS (BY SIMILARITY).			
	CC	- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC AND			
	CC	VARIES AMONG DIFFERENT ALLELES.			
	CC	- SIMILARITY: THE N-TERMINAL DOMAIN SHOWS SOME SIMILARITY TO THAT			
	CC	OF SILKWORM HEMOCYTIN.			
	CC	- SIMILARITY: CONTAINS 2 VWFC DOMAINS.			
	CC	- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTEINE KNOT-LIKE DOMAIN (CTCK).			

Search completed: April 26, 2002, 16:32:20
Job time: 735 sec

	Query Match	8.8%	Score 643.5;	DB 2;	Length 990;
	Best Local Similarity	24.2%;	Pred. No. 2e-20;		
	Matches 246;	Conservative 148;	Mismatches 424;	Indels 199;	Gaps 40;
QY	74	SPPSKKAPPGSG-----ASTFIKSTKRSPKNKKKT-----KKVIESEITEE	119		
		: :	:	:	:
Db	57	SPDAKKRPANGLPUPPKKAKESSSDSDEEPKAKRAOAGGKPPVKAQPCKA	116		
		: :	:	:	:
QY	120	HVSNEGSSSSSSSSSSSTIWK-IKSSNSAAANRELOKKLVKDNNKNT-----KK	172		
		: :	:	:	:
Db	117	KSSSEDSDSDSEETKPPAKRPAOTPPVAAVKTTQKAKSSSESSEADEASKKK	176		
		: :	:	:	:
QY	173	KPTPKPP---VVDAGSLONGDFKVTTDTSTTQHKNKVTSPKITTAKPINRPSLPPN	229		
		: :	:	:	:

```

RESULT. 15
T11622
    extensin class 1 precursor - cowpea
C:Species: Vigna unguiculata (cowpea)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 17-Nov-2000
C:Accession: T11622; S54155
R:Arsenijevic-Maksimovic, I.; Broughton, W.J.; Krause, A.
Mol. Plant Microbe Interact. 10, 95-101, 1997
A:Title: Rhicobia modulate root-hair-specific expression of extensin genes.
A:Reference number: Z17301; MUID:97155574
A:Accession: T11622
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA

```

QY	478	----	-----PPTPKPEAPTTTTKSAPTTT--KEPAPTTT	504
Db	2286	GLVCRNREQVKMCFNVEIRVFCNCYHCPSIPATSSATSSPTGGTWILTKLTUTA	2345	
QY	505	KSAPTTPKPSPTTKKEPAPTTPKKAPPTTKEPAPTTKEPAPTTKPKAPTTTKKPA-	563	
Db	2346	TTTESTGSTATSSTOGPPAGTPH--VSTTATTTTPTVSSK-----ATPSSPGTATALPAL	2399	
QY	564	-----PTAPKEPA-----PPTPKETAPTPKKLTP-----	588	
Db	2400	RSTATPTATSFATPSSSLGTTWTRLSTQTTTPMATMSTATSPETVHTSVLTITAT	2459	
QY	589	-----TTPKELAPTTPEKAPTT-----PEELAPTTPEEP-----TPTTPEEPAP	628	
Db	2460	TTGATGVAIPSSPTPGTAHTTKVTTTTTGTFTVPSSSPGTARPPVWISTTTPTTSGS	2519	
QY	629	TPPKAAAPNTPKEPAP--TPPKEPAPT--TPKEAPNT--TPKE--TAPTPPKGTAPTTL	679	
Db	2520	TVTPSSIPGTHPTVLTWTTTQPVATGSMATPSSSTQTSQTPPSSLITAITVIATGTT-	2578	
QY	680	KEPAPTPPKPAKELAPT--TKTEPTS-----TTSOKPAPTTPKGT--APT	723	
Db	2579	NPSTGCTPIPELTTTATTPAATSSVTVPSSALGTHTPPVNTTATTHGRSLSPSS	2637	
QY	724	P-----KEPAPTTPKEPAPTTPKGTAPT--TLKEPAP--TTPKKP	759	
Db	2638	PHTVRTAMTSATGTLGTHITPEPTSGTSHTPAAITGTTTTTSPALSSPHSPSTTESPP	2697	
QY	760	APKELAPTTTKGPTSTSDKPAPTPKETAPT--PKEP-----	796	
Db	2698	SP---GTTTPGHTATSRITATATSKTRTSTLLPSQPTSAPIITVVVTGCEPQCAWSE	2753	
QY	797	-----	796	
Db	2754	WLDYSYMPGSGGDEFTYSNIRAAAGVACEOPGLECRATAQCVPIGELQGVCESLD	2813	
QY	797	-----APTTPKPAPTTPTPTPTT---SEVSTP	822	
Db	2814	FGLVCRNREQVKMCFNVEIRVFCNCYHCPSIPATSSATPSSPTGGTWILTELT	2873	
QY	823	TTTKETPTIHKSPDESTPELSAETPKALENSKPEGVGTTK-----TP	866	
Db	2874	ATTTASTGSTATP--SSTP--GTAPPPKVLTSPTATPTATSSKATSSSPRTATPLVLT	2930	
QY	867	AAK-----PEMTTAKDKTTEROLRTPETT--TAAPKMTKETAT	905	
Db	2931	TATKATATSVPIPSSTLGTGILPEQTTTPVATMSTIHPSSPTETHSTVLTTKATR	2990	
QY	906	TTEKT-----TESKITATQVTS--TTQDTRP-----PKITTLKTTILAP	945	
Db	2991	ATSSSTSPSTPGTWILTELTAAATTTAGTGATPSPSTPGTTWILTELTATTAFTAST	3050	
QY	946	KVTTT--KKTITITTEIMNKPPEE---TAKPKDRAT--NSKATTPKO-----KPTKAPKK	992	
Db	3051	GSTATLSSPTGGTWILTELPSTTAVTPADGSTATASQATACTPHVSTTATTPVTSSK	3110	
QY	993	PTSTKKPKT--MPRVRRKPTTTPRKM-----ST	1020	
Db	3111	ATPSSSPGTATPALRSTATTPTATSFATIPSSSLGTTWTRLSTQTTPTATMTSTATPSS	3170	
QY	1021	MPELNPTSRIAEMLOTT-----TRNQNPNSKLVEVNPKSEDA--CAEGETPHMLLR	1072	
Db	3171	TPETVHTSVLTATTTAGTGSVATPSSPTGTAHTTKVPTTTTGTATPSSSGTALT	3230	
QY	1073	PHVMEPVTMDYLPVRPNQGLIINPMLSDETNICNGKPVDOGTLATLRNGTLVAFRGHYF	1132	
Db	3231	PPVWISTTTPITTP--TTSGSTVTPSSIPGTHTARVLTTTITTTTATGSM-----	3280	
QY	1133	WMLSPSPSPARRITEVWGPSPIDIVFTRCNEGT	1170	
Db	3281	-----ATPSS---TOTSGTPPSLTATTTATGST	3309	

RESULT 13

T34513
hypothetical protein ZK783.1 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34513
R:Favello, A.; Vaudin, M.
submitted to the EMBL Data Library, August 1994
A:Description: The sequence of *C. elegans* cosmid ZK783.
A:Reference number: Z21536
A:Accession: T34513
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-3507 <FAV>
A:Cross-references: EMBL:U13646; PTDN:AAC24418.1; GSPDB:GN00021; CESP:ZK783.1
A:Experimental source: strain Bristol N2; clone ZK783
C:Genetics:
A:Gene: CESP:ZK783.1
A:Map position: 3
A:Introns: 14/1; 48/2; 84/1; 196/3; 303/1; 381/1; 586/1; 605/1; 1175/3; 1207/1; 1409/3
3504/1

Query Match

Best Local Similarity 23.3%; Pred. No. 4.9e-20;

Conservative	Mismatches	Indels	Gaps
283	182	253	52

Query Match	8.9%;	Score 649;	DB 2;	Length 3570;
Best Local similarity	22.5%;	Pred. No. 4.3e-20;		
Matches 373;	Conservative 111;	Mismatches 584;	Indels 590;	Gaps
QY	62	RVCTAAEVHNPTSPSSKKA---PPSGASQTIKSTTKRSPKPPKNNKTKKVVIESEETEE	119	
Db	1693	RTATTLPLVLTATKSTATFTPIPSSTLGTTGTSQNRPPHPMATMSIHPSSTPETHT	1752	
QY	120	HSV-----SENOESSSSSSSSSTIWKIKSSKNSAANRELQKLKVKDNKNKRT--KK	172	
Db	1753	STVLTKATTTATSSMSPSTPGTWT-ILTELTAATTAALPHGTPSPSTPGTTWILT	1811	
QY	173	KPNPKPVVDEAGS-----GLDNGDFKV-----TTPDSTTQHKNKVSPTKTTAKPI	220	
Db	1812	EPSTATVTPGTSATASSTRATAGTLKVLSTATTPT-TVISSRATPSSSGTATA--	1867	
QY	221	NRPSPPPNSDTSKETSUTVNKETVETL-----KETTINKQTSFDGKEKTSKAKETQSI	275	
Db	1868	--LPALRSTATTPTATSVTAIPSSSLGTAWRLSQTTPATMTATNPSTPETHVSTV	1925	
QY	276	EKTSAKDLAPTCKVIAPKTPKAEYTK-----GPAITTP-----KEPTP	314	
Db	1926	LTTATTATRTGSAVPSPSTPGTAHTTKVPTTTTGTATPSSSGTALTPPVWLSTTTP	1985	
QY	315	T-----TPKE-----PASTTPKEPTTTTKSAPTPPKP	343	
Db	1986	TTRGSTVTPSSIPGTHHTATVLTTTITVATGSMATPSSSTSGTPPSLTATTATAT	2045	
QY	344	ATTTKSAKAP---TTPKEPAPTTT-KEPAPT-----TPKEPAPTTTKEPAPTTKSA-----	390	
Db	2046	GSTTNPSPSTPGTTPPPVLLTTTATTPAATSVTPSSALGTTHTPPVPNTATTHGRSLP	2105	
QY	391	PTTP-----KEPAPTPKKAPTT-----PKEPAPTPPKPAPTT	426	
Db	2106	PSSPHVTATATSATGSLGTHHTERTSTGTSHTPPAATCTGTQSPALUSSHPSPSRITTE	2165	
QY	427	KEPAPTTKEPAPT--TPKEPAPTAPKAPAPTTPKEPAPTT-PKEPAPTTTPKPS-----	477	
Db	2166	SPSPSGTTPGHTGTSRTTATATPASKTRSTLLPSSPTSAPITVTVVTTCCEPCANSEW	2225	
QY	478	-----	477	
Db	2226	LDYSVPMPPSGGDFDVTYNIRAAAGVACBQPLGLECRAQAGQVPLRELIGQWVECSLDF	2285	

Mon Apr 29 08:35:21 2002

Biochem. Biophys. Res. Commun. 165, 644-649, 1989
A:Title: Sequence analysis of the 5' region of the human DF3 breast carcinoma-associated
A:Reference number: A36735; MUID:90088473
A:Accession: A36735
A:Molecule type: mRNA
A:Residues: 1-142, 'Q', '144-162, 'Q', '164-168 <ABE>
A:Cross-references: EMBL:M31823; NID:g181542; PIDN:AAA35757.1; PID:g181543
R:Masuzawa, Y.; Miyauchi, T.; Hamanoue, M.; Ando, S.; Yoshida, J.; Takao, S.; Shimazu, H.
J. Biochem. 112, 609-615, 1992
A:Title: A novel core protein as well as polymorphic epithelial mucin carry peanut agglu
A:Reference number: JX0235; MUID:93123189
A:Accession: PX0066
A:Molecule type: mRNA
A:Residues: 998-1011, 'ES', '1014-1017; 1018-1032, 'T', '1034-1037; 1038-1057 <MAS>
A:Experimental source: gastric carcinoma cell
R:Rihlan-Licht, S.; Baruch, A.; Elroy-Stein, O.; Keydar, I.; Wreschner, D.H.
FEBS Lett. 336, 130-136, 1994
A:Title: Tyrosine phosphorylation of the MUC1 breast cancer membrane proteins cytokine
A:Reference number: S51026; MUID:95080414
A:Contents: annotation
A:Note: undetermined tyrosine residues in the carboxyl-terminal non-repetitive region ar
C:Comment: This protein is length polymorphic. Individuals may have between 21 and 125 c
partial repeats. The repeat shown is defined by SmaI nuclease sites.
C:Comment: Serine and threonine residues in the tandem repeat domain are extensively gly
C:Comment: For an alternative splice form without a tandem repeat domain, see PIR:S48146
C:Genetics:
A:Gene: GDB:MUC1; PUM
A:Cross-references: GDB:120705; OMIM:158340
A:Map position: 1q21-q23
A:Introns: 20/1; 62/3; 1165/3; 1184/2; 1230/1; 1270/3; 1320/3
C:Superfamily: polymorphic epithelial mucin
C:Keywords: alternative splicing; duplication; glycoprotein; glycoprotein; polymorphis
F:1-1344/Product: mucin 1 precursor, splice form A #status predicted <PREA>
F:1-62/Region: mucin 1 amino-terminal non-repetitive
F:1-23/Domain: signal sequence #link PREA #status predicted <SIGA>
F:1-19, 29-32/Domain: signal sequence #link PREB #status predicted <SIGB>
F:1-19, 29-32/Domain: mucin 1 precursor, splice form B #status predicted <PREB>
F:1-19, 29-32/Domain: mucin 1 precursor, splice form C #status predicted <PREC>
F:1-19, 29-32/Domain: mucin 1 precursor, splice form D #status predicted <PRED>
F:138-1017/Region: 20-residue repeats (GSTAPAHGVTSAPDTRAP)
F:1143-1344/Region: mucin 1 carboxyl-terminal non-repetitive
F:1245-1272/Domain: transmembrane #status predicted <TRM>
F:1046, 1064, 1118, 1144, 1222/Binding site: carboxylate (Asn) (covalent) #status predicted
F:1213/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 9.2%; Score 672; DB 1; Length 1344;
Best Local Similarity 27.8%; Pred. No. 1.7e-21;
Matches 310; Conservative 78; Mismatches 472; Indels 256; Gaps 56;

QY 238 LTVNKKETVETKETTNNK--OTSDGKEKTSKETSIEKTSKAD-LAPTSKVLAKPT 294
Db 15 LTVLTATTAPPAIVTVGSGHASTPGGKETTSATQSSVPSSTKNAVMSVLSHS 74
QY 295 P-KAETTTKGP--ALTTPKEP-----TPTTPKEPASTTP-----KE 327
Db 75 PGSGSSTTQGDVTLAPATEPASGSAATWGDVTSVPVTRPALGSTTPPAHDVTSAPDNK 134
QY 328 PTPPT-----IKSAPTTKPEAPTTT-----KSAPTTKPEAPTTTKEP-----A 367
Db 135 PAPSSTAPPAHGVTSAPDT--RPAGSTAPPAHGVTSAPDT--RPAGSTAPPAHGVTS 190
QY 368 PTTTPKEAPTTTKEPAPTTTTSAPTTTPKEAPTTT-----KKPAPTTTKEP----- 413
Db 191 PDT--RPAGSTAPPAHGVTSAPDT--RPAGSTAPPAHGVTSAPDT--RPAGSTAPPAHGV 247
QY 414 --APTTPKEPTTTP-----KEPAPTTKEP-----APTTPKEAPTTAPK---- 450
Db 248 TSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGV 307
QY 451 KPAPTTKPEAPTTTPK----EPAPTTTKEPPTTTPKEAPTTT--TKSAPTTTKEPAPTT 503
Db 308 TSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGV 367

QY 504 TKSAPTTKPEPSPTTTKEP-----APTTPKEAPTTTPK-----KPAPTTKPEAPTTTPK- 552
Db 368 T-SAPDT--RPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPA 424
QY 553 ---EPAPTTTTPKAP 608
Db 425 HGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPA 484
QY 609 ---ELAPTTPEPTTTP-----EEPAP--TTPKA----AAPNTKPEAPTTTPK- 648
Db 485 HGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPA 544
QY 649 ---EPAPTTTPKEAPTTTPKAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 695
Db 545 HGVTSAPDTRPAGSTAP--PAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGST 600
QY 696 APTTTKEPTTTSAPAP--TTPKGTAPTTTPKEAPTTTPKEAPTTTPK-----TAPTTLK 749
Db 601 AP-PAHGVTSAPDTRPAGSTAPPAHGVTS-----APDTRPAGSTAPPAHGVTSAPDTR 655
QY 750 EPAPTT-----KKPAPKELAPTTTKGTSTTSAPAP--TTPK-----ETAPTTTP 793
Db 656 APGSTAPPAHGVTSAPDTRPAGSTAP--PAHGVTSAPDTRPAGSTAPPAHGVTSAPDTR 714
QY 794 KEAPTTTP-----KKPAPTTTPETPTTSEVSTPTTTPKEPT-----IHKSPD 836
Db 715 PAPSSTAPPAHGVTSAPDTRPAGS--TAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPD 772
QY 837 ESTPELSAPTPKALENSPKPGVTTKTPAA-----TKPEMTTTAKDK---TTERD 885
Db 773 TRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPD 832
QY 886 LRTTPPTT-----TAAKMTKETATTTTEKTESKITATTQTQVSTTTQDTPP--FKIT 936
Db 833 TRPAGSTAPPAHGVTSAP--DTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGV 888
QY 937 TLKTTTLAPKVVTTTKTTTITTT--EIMNKPBEAKPKDRATNSKATTPKQKPKPKPKPT 994
Db 889 SAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAG--GSTAPPAH 946
QY 995 STKKPKTMRVRKPKTTTPTPKMTSTMPENP---TSRIAEAMLOT--TTPR---NOTPN 1046
Db 947 VTSAPDTRP--APGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPP 1003
QY 1047 SKLVEVNPKSEDAGAGETPHMLLRPHVEMPEVTPDMYLPVNPVNOGIIINPMLSDTN 1106
Db 1004 AHGVTSAPDTRPAGSTAPPAH-----GVTSAPDNRPALGSTA----PPVHNVT 1049
QY 1107 ICNGKPVDDGLTTLRNGTLVAFRGHYFWMLSPFSPPS 1142
Db 1050 ASGSASGASTLVHNGTSARATTTTPASKSTPFSIPS 1085

RESULT 11
I38346
elastic titin - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 21-Jul-2000
C:Accession: I38346
R:Label: S.; Kolmerer, B.
Science 270, 293-296, 1995
A:Title: Titins: giant proteins in charge of muscle ultrastructure and elasticity.
A:Reference number: A57430; MUID:96026330
A:Accession: I38346
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-7962 <RES>
A:Cross-references: EMBL:X90569; NID:g1017426; PIDN:CAA62189.1; PID:g1017427
C:Genetics:
A:Gene: GDB:TTN
A:Cross-references: GDB:127867; OMIM:188840
A:Map position: 2q31-2q31

QY	725	-KEPAPTPKAPAPTPKGTATTTKEAPATTPKPKAPKELAPATTTKGTSTTSDKAPT	783
DB	1001	MSPPPPPEVKSPPPAPVSSPPPVKSPPPAPVSSPP---PPVKSPPPPAPVSSPPPV	1057
QY	784	-TPKETATTPKAPATTPKPKAPATTPPTTSEVSTPTTKEPTTIHKSPPDESTEEL	842
DB	1058	KSPPPAPISPPPPPVKSPPPAPVSS---SPPPPVKSPPPAPVSSPPPIKSPPPAPVS	1115
QY	843	SAEPTPKALENSPKPGVYTTTKTPAATKEMTTAKDKTTERDLRTTPTET	892
DB	1116	SPPPAP-----VKPPSLP-PPAPVSSPPPVVTPAPPKKEQSLPPPAES	1158
RESULT 8			
T18535			
high molecular mass nuclear antigen - chicken (fragment)			
C:Species: Gallus gallus (chicken)			
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999			
C:Accession: T18535			
R:Shimada, K.; Harata, M.; Mizuno, S.			
J. Cell Sci. 110, 3031-3041, 1997			
A:Title: A nuclear matrix-associated high molecular mass nuclear antigen, HMNA, of chick			
A:Reference number: Z18955; MUID:9803440			
A:Accession: T18535			
A:Status: preliminary; translated from GB/EMBL/DBJ			
A:Molecule type: mRNA			
A:Residues: 1-1151 <SHI>			
A:Cross-references: EMBL:D88440; NID:d1177138; PID:d1025045; PIDN:BAA24137.1			
Query Match 9.5%; Score 695; DB 2: Length 1151;			
Best Local Similarity 25.3%; Pred. No. 1.6e-22;			
Matches 304; Conservative 108; Mismatches 466; Indels 324; Gaps 59;			
QY	67	AEVHNPTSPSSKK-----APPGSASQTIKSTTKRSPKPPNKKTKKIVIEEITEEH	120
DB	39	AELH-PAAPQPPKVVIGAPPPTGTEPT-----PPSKPTDGADAAKASAEILTSPP	90
QY	121	SVSNQESSSSSSSSSSSTIWKTKSKNSAANRELOKLVKDNKNRKTTPKPPV	180
DB	91	PASPPDPGPKAPSGAGAEAGTPPPSQGPAG-----TTPP--	126
QY	181	VDEAGSLDNGDFKVTPTDSTTQHNKVSST--PKIITAKPINRPSLPPNSDTSKETSL	238
DB	127	-SQGAAGAPKGDGVAQPSGKSGADGPKAAGDVFKATA-ATEARP-----ASAASP	176
QY	239	TVNKETTVEKRETTTNNK--QTSTDGKEKTTSAKETOSIEKTSAKDLAPTSKVLAKTPK	296
DB	177	TVPRKATAEATAVTAASQSAPKAAATAAAVTA--SQSAPKATY-EVKPAAAVAKEKA	232
QY	297	AETTTKGALTPKPTP-TTPKPEASTTPKEP---TPTTIKSAPTTKPEAPTT-----	347
DB	233	VTAAAAAPKATAEAKAPAPVTSPTIPCSAEAKPLTAASPTASKA--TAAKAPVATASLM	290
QY	348	-----TKSAPTTKPEAPTTK-----EPAPTTKPEAPT	377
DB	291	ATKVTAEAKPAPSPVKPRTTDTTAVTATAPKAGPDVKPVAVCAEAKAPAPPPPOQLPK	350
QY	378	TTKEAPATTTKSAPT-----PKPA-----PTTKKAPATTPKPEAPTTKPTPT	424
DB	351	AAAAAAPTGTCLKATAPPHGSPRANSHVTVTPPNVPRAAAATVP--TAGAVPKASTGT	408
QY	425	T-----PKPEAPTTKPEAPTTKPEAPATPKKAPATTPKPEAPTTKPEP---APTTFEKS	477
DB	409	TPAAAPQOQVPP---KAAPVTPSPQOAVPRATAAA---APVTQOPVTKAAATTNATP	461
QY	478	PTPKPEAPTTTKSAPTTTKPEAP---TTTKSAPTTTKPEPSPTTTKEP---APTTPKEP	530
DB	462	PPQIPKAAATTTATVPTVTPQOIPKAGTDAAPPAPVAPKPSDGRAATGVPNAAATDPOKP	521
QY	531	APTTPKKAPATTPKPEAP-----TTPKPEAPT-TTKKAPATPK-----EPAPT	573

Query Match 9.3%; Score 680; DB 2: Length 1229;
Best Local Similarity 26.6%; Pred. No. 7.3e-22;
Matches 334; Conservative 105; Mismatches 394; Indels 422; Gaps 60;

A:Cross-references: EMBL:U48363; NID:g1666688; PID:g1666689; PIDN:AAB18732.1

C:Genetics:

A:Map position: 10

A:Introns: 24/1; 1996/1; 2024/3; 2050/3; 2099/3; 2142/3; 2183/3

A:Note: differential splicing converts alphaNAC into a tissue-specific DNA-binding active

C:keywords: alternative splicing; DNA binding; transcription factor

Query Match 11.1%; Score 812; DB 2; Length 2187;
Best Local Similarity 27.4%; Pred. No. 3.4e-27;
Matches 311; Conservative 128; Mismatches 449; Indels 246; Gaps 51;

QY 71 NPTSPSSKAPPPSGASOTIKSTKRSKPPNKKKTKKVBIESEBITEHSEHSSENQESS 130

DB 746 SPT-PPSSKGAVPVPGA-----PPSPKGAIVPTSEISSKQVPAEILPSQ 792

QY 131 SSSSSSSSTIWKIKSSKNSAANRELQKLVKDKNKKTKKTPKPPVVDAGSLDN 190

DB 793 KTEPVTAASRLISAVQSPKVDPIMSD-----VTPTSPKKTSATAVPK----- 833

QY 191 GDPKVTPTDSTTQHNK-----VSTSPKITTAKPINRPSLPPNSDTSKETSLS--TVNKE 243

DB 834 -----DTSATLSKSVPAVTSLEPPKAPVAPSNEATIVPEIPTSLSKNALAAATPK 885

QY 244 T-----TVETKETTTTNTKOTSDGKEKTTSAKETOSIEKTSKADLAPTSKVLAKTPKA 297

DB 886 TLATSIPKVTSPSPKTPKSVLSKGAAMTSKAT---ETAASKDVPSQ--FPKREPLL 940

QY 298 ETTTKGALTTKREPTTTPKEPASITPKETPTTIKSAPTTPKE-PAPTPTTKSATPTPK 356

DB 941 QHV---PPTSPKSPVSDTLGSLATSPPKGPPATIAETTYPKPKPAASKTTATPS 997

QY 357 -----EPAPTTPKEAPTTPKEAPTTPKEAPTTP--TKSAPT----- 392

DB 998 PEGTAVPLEIPPCSKKAPKTAAPKSSATSSSKRAPKTAVSKEIPSQVTAVPLEISLP 1057

QY 393 -----TPKEAPTTPPK--PAPTTPKEAPTTPKEPTTTPKEAPTTPKEAPTTPKE 443

DB 1058 LKETSATPGEKSASSPKSPKTAGKE---TPPGVTAVPPEISLPKPKETQONATPNE 1114

QY 444 PAPTAPKAPATT--PKEPAP-----TTPKEAPTTPKEPSPTPK-----EPAPTTPKSA 492

DB 1115 SLAASSQKRSKTSVPKETPPGGVTAMPELIPSAPOKAPKTAVPKQIPTPEDAVTILAGS 1174

QY 493 PTTTKEAPTTP--KSAPTTPK-----EPSPTTKEAPTTPKEAPTTPKPKAPATT- 542

DB 1175 PLSPKKASKATAAPKEAPTSPGVIAVSGEISPSPKTSKTAAPKENSATLPPKRSPKTA 1234

QY 543 -PKE-----PAPTTPKEAP-----TTTKKAPATAPKEAPTTPKETAPT 582

DB 1235 APKETPATSSEGVTAVPSEISPSPTPASKGVPVTLTPKAPNALAE-SPASPKKVPKTA 1293

QY 583 PKKLTPTTPEKLAPTTPEK-PAPTTPPEELAPTTPEPTPT-----PEEPAPTTPKA 633

DB 1294 APEETSTP-----SPQIKPVAGPKASATPPSKKTPKTAAPKETSAPSEGVTAVDLE 1347

QY 634 AAPNTKEAPTTPKE-PAPTTPKKEAPTTPKETAPTTPKGTAPTTLKEAPTTP----- 687

DB 1348 IPPSPRKAPKTAAPKETPAPS--PEGATTAPVQIIPSPKPKSKKAGSKE-TPTTSPSEGV 1404

QY 688 -----KKPAPKELAPTTTKEPTS-----TTSDKPAPTTPKGTATTPKE 726

DB 1405 TAAPLEIPISSKTSKMAKPELTVTPSSKKLSQTVGPKETSLEGATAVPLEIIPSHKKA 1464

QY 727 PAPTTPKE-PAPTTPKGTATTLKEAPTTPKPAKELAPTTKGTSTTSDKPAPTTP 785

DB 1465 PKTVDPKQVPLTSPK-DAPTTLAE-SPSPKK-APKTAAPSESR-VTVVPEKPA-TPQ 1519

QY 786 KETAPTTPKEAPTTPKPKAPATTPEP-----PPTTSEVSTPTTK-----EPTTIHKS 834

DB 1520 KASGTTASKVPVPAETQEVAVSSREIFVTPAVPPVKNPSSHKTSKTIELKEAPATLPSS 1579

QY 835 PDESTPELSAETPKALENSPKPE-PGVPTTKTPAATKPEMTTAKDKTTERDLRTPTPT 893

DB 1580 PTKSPKIPSSKAPRI--SAPKEFPASPSIK-PVIT--SLAQTAAPPSSLOKAPSTTIPKEN 1634

QY 894 TAAPKMTKETATTTEKTSKTTATTTOVTSITTQDTTTPFKITTLTKTTTLAPKVTTTKKT 953

DB 1635 LAAPAV---LPVSSSPAAPARASLASPATAAPOTAPKREATTTPSCKKAATETPIET 1690

QY 954 ITTEIMNKPEETAKPKDRATNSKATTPKPKTPKAPKPTSTTKPKTMPRVKPKPTTPT 1013

DB 1691 STAPSLGAPKPESE-----TSVSKVLMSSP-----PKRASSSKRASTLP----- 1730

QY 1014 PRKMTSTMPLEINTPSRTIAEAMLTQTRPNQTPNSKLVEVNPKESEDAGGAGETP 1067

DB 1731 ----ATTPLSLKEASVLS-----PTATSSGKDSHISVPS-DACSTGTTTP 1770

RESULT 6

S48478 glukan 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast (Saccharomyces cerevisiae)

N:Alternate names: extracellular glucosylase; mucin-like protein MUC1; protein YIR01

C:Species: Saccharomyces cerevisiae

C:Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text_change 12-Nov-1999

C:Accession: S48478; A26877; B26877; S27281; JG6123

R:Rowley, K.

submitted to the EMBL Data Library, October 1994

A:Reference number: S48478

A:Molecule type: DNA

A:Residues: 1-1367 <ROW>

A:Cross-references: GB:247047; EMBL:Z38061; NID:g603997; PID:g763364; GSPDB:GN000009;

R:Yamashita, I.; Nakamura, M.; Fukui, S.

J. Bacteriol. 169, 2142-2149, 1987

A:Title: Gene fusion is a possible mechanism underlying the evolution of STAL.

A:Reference number: A91831; MUID:87194600

A:Accession: A26877

A:Molecule type: DNA

A:Residues: 1-242 <XNA>

A:Cross-references: EMBL:M16164; NID:g172522; PIDN:AAA35014.1; PID:g172525

A:Accession: B26877

A:Molecule type: DNA

A:Residues: 762-1331 <YA2>

A:Cross-references: EMBL:M16165; NID:g172523; PIDN:AAA35015.1; PID:g172526

R:Pardo, J.M.; Ianez, E.; Zalacain, M.; Claros, M.G.; Jimenez, A.

FEBS Lett. 239, 179-184, 1988

A:Title: Similar short elements in the 5' regions of the STA2 and SGA genes from Sacc

A:Reference number: S27281; MUID:89031230

A:Accession: S27281

A:Molecule type: DNA

A:Residues: 1-31 <PAR>

A:Cross-references: EMBL:X13857; NID:g4551; PIDN:CAA32069.1; PID:g4552

R:Lambrechts, M.G.; Bauer, F.F.; Marmur, J.; Pretorius, I.S.

Proc. Natl. Acad. Sci. U.S.A. 93, 8419-8424, 1996

A:Title: Muc1, a mucin-like protein that is regulated by Msi10, is critical for pseud

A:Reference number: JC6123; MUID:96323237

A:Accession: JC6123

A:Status: nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 1-1367 <LAW>

A:Cross-references: GB:U30626; NID:g1304386; PIDN:AAC49609.1; PID:g1304387

C:Genetics:

A:Gene: SGD:MUC1; STA2; MAL5; DEX2; SGD:S0001458

A:Cross-references: MIPS:YIR019C; SGD:S0001458

A:Map position: 9R

C:Superfamily: yeast glukan 1,4-alpha-glucosidase homolog; glukan 1,4-alpha-glucosida

C:Keywords: glycosidase; hydrolase; polysaccharide degradation; transmembrane protein

F:5-21/Domain: transmembrane #status predicted <TM1>

F:1350-1366/Domain: transmembrane #status predicted <TM2>

Query Match 11.1%; Score 810.5; DB 1; Length 1367;

Best Local Similarity 27.2%; Pred. No. 2.5e-27;

Matches 332; Conservative 114; Mismatches 519; Indels 255; Gaps 54;

Db	1048	ASTEETTYAPTEETTYAPAEETVPYEETTYAPTEETTYAPTEETTYAPTEETTYAPTE	1107
QY	783	TTPKETAPTTPKEAPITPKAPATTPETPPPTSEVSTPTTKKEPT--TIHKSPDESTP	840
Db	1108	ETTYAPAEETVPYEETTYAPTEETTYAPTEETMYAPTEETTYGTEETTY-APTEATT	1166
QY	841	ELSAEPTPKA-LENSKPEPG-----VPTTKTPAATKPKMTTAKOKTTDERLRTTP---	890
Db	1167	YAPTEETPYAPTEETTYEPTGETTYAPTEETTYAPTEETTYAPTEETTYAPTEETPYEPT	1226
QY	891	ENTTAAPKMTETATTTEKTESKIRATTQVSTTTQDPT-----PFKITTY---	937
Db	1227	EEITYAP-----TEETTYEPTTYAPTEETTYAPTEETTYAPTEETMYAPIDEITYGP	1281
QY	938	LKTTTLAPKVTTTKTITTTIEMNKPEETAKPKDRAT--NSKATTPKP-QKPTKAPKKPT	994
Db	1282	TEETTYAPTEATTYAPTEETPYAPTEETTYEPTGETTYAPTEETTYAPTEETTYAPMEET	1341
QY	995	-----STKKP-KTMPVRVRPKPTTPPKMTSTMPLELNPTSRIAEMLQITTRP	1041
Db	1342	PYPAEESTSTVSTEKPCNTEETDEPTDEPTDE--PSEDEPTDEPTDLPTDEFPSTP	1399
QY	1042	--NOTPNSKLVENPKSEDAG	1060
Db	1400	CDNOGINGIGVENKVRNNAG	1420

```

RESULT      4
hypothetical protein F35A5.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 18-Feb-2000
C:Accession: T16251
R:Leimbach, D.
submitted to the EMBL Data Library, January 1996
A:Description: The sequence of C. elegans cosmid F35A5.
A:Reference number: Z18485
A:Accession: T16251
A:Status: preliminary; translated from GH/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1774 <LEI>
A:Cross-references: EMBL:U46675; NID:g1166613; PID:g1166621; PIDN:AAB52641.1
A:Experimental source: strain Bristol N2; clone F35A5
C:Genetics:
A:Gene: CESP:F35A5.1
A:Map position: X
A:Introns: 1272/2

```

Query Match 11.68; Score 846; DB 2; Length 1274;
Best Local Similarity 27.63; Pred. No. 7.2e+29;
Matches 313; Conservative 119; Mismatches 455; Indels 248; Gaps 60;

QY	76	PSSKKAAPPSSGASQTIKSTTKRSP----	KPP--NKKTKKVISEEITTEHVSVENOES	128
		:	:	
Db	185	PSPKAAKSKEHDIVPPTPIKNPAKKWPPEDE	VETIEIKPEPATRKVPALAKKEP	244
		: :	:	
QY	129	SSSSSSSSSSTIWKIKSSNSAANRELQKLKVKD	N-KNRKK-KP-----TP-----	176
		:	:	
Db	245	STSVPYSDDSPKKV-----PVKKEVEVPTT	IKNPTKKWKPPEDETPTVEEVK	294
		:	:	
QY	177	KPPVVDENGSLONGD-----FKVTPTDTSTT	OHNKVSTSPKITTAKEPI-NPRFSLPNSD	231
		: :	:	
Db	295	EPPVPEKKAPVLKKKOPAPAAROPSKAAPK	KVEPSSPVVPTPVKNPVKKYPWE	354
		:	:	
QY	232	TSKETSUTVAKETTVETKETTNTKTSTDGREK	TTSAKETOSEKTSAKDLAPTSKVLA	291
		:	:	
Db	355	VDDSPAEEAVKPSAPEKKTPLVRKREPSSST	TSSDPSPKAAPKAVKPDSSPKKATPL	414
		:	:	
QY	292	KPTPKAE-----TTTTGPA-----	LTTPKEPTPTP-----KEPASATTPK	326
		:	:	
Db	415	QADPKAEVPTPVKNPVKKYKXPPEVDDDF	VEEVKQEPAPAKKTPVLKRKEPAKOTA	474

QY	327	EP----	TP-TTIKSAPRTTPEKAPTPTTKSAPRTTPEKAPTPTTKBPA-----	367		
Db	475	KPATSKPTETPEKDPVKPRDSSPKVAAKPOSAQAPA--TPVKNPVKKWRPWEDEDTPA	533			
QY	368	-----PITTPKEAPRTTTKEAPRT-----	TTKSAPRTTPKEB-----	APTTPPKAPRTTP	410	
Db	534	DDVSKPDAANTPSLAKKDOPAPAKESLKPKADTKAPAKPRDPSPKKVAPTAPEKKTVPVA	593			
QY	411	KE-----	PAPRTTPEKPTPTTPKBPARTTPKBPARTTPK--EPAETAOKKAP-----	TTPKE	459	
Db	594	KKEPAGADSKTKEPEKSKRDPSPKKAHPAKVPVKETVAVAAVVKKPEIPISKPKDTAPKK	653			
QY	460	PAPRTTPEKAPTPTTKBEP-----	SPTTPKBPARTTTSKAPTPTTTTKEBAPTPTTKSAPTTP--KEP	514		
Db	654	AEPNSPVVP--PTBVKNPVKKWPPWEDDDAKPAKPVSLPEPEKK--TPVLAKKATPKPDSEA	711			
QY	515	SPTTTTKEBAPTTPK--EAPARTTPPKAP--	TTPEKAPTTPKBPARTTPTTKKPAPTA--PKE	569		
Db	712	AADPVSGSSKOPKLAKKAPVKPRDPSPMKAVPIKPAKPT--EVPPAVKKKPEPVAKS RD	769			
QY	570	PAPRTTPEKAPTTPKKLTPTTP-----	BKLAPRTTPEKAPTTPBELAPRTTPEB-----	T	619	
Db	770	PSPKKAH-AEPNSP--VVPPTPVKNPVKKWPPWEDDDAEPVNVNPEBEKKTTPVLAKKT	826			
QY	620	PTTPEEPAPRTTPKAAAPNTPEKAPTTPKBPARTT-PKEBAPTTPKETAPRTTPKGTAPRTT	678			
Db	827	PVKPRDPSKKAHPAKSTGTDAPSVSVKKPEPVSKPEPSKPBAPNSPVVP---PTP	882			
QY	679	LKEBAPTTPKKPAPKELAPT--TTKEPTSTTSKAPRTTPKGATTPKBPARTTPKBPARTTPKEAP	737			
Db	883	VKNPVKKW-KPWEDDEPTEEVKKSE--PEKKTTPVLAK-KEPEKPKO-APKVAAPKRD	937			
QY	738	TPPKGATPTTLKEAPRT-----	TPKKAP-----	KELAPTTTKGPTS-----	TTSDK	779
Db	938	PSFKKAVE--KEPAKVAARKPRDLSPSKAIPIFANTQEAPPTPVKNPVKKWKPPWEDDDE	995			
QY	780	P-----	APTTPKET-----	APTTPKEBAPTTPKBPARTTPEPTTPBTSEVSTPTTTKEP	828	
Db	996	PAEPSVAPEBEKKTTPVLAKKAPAKPRD--	SPKKAAPVAAK--	PDPKTPEV--PPTPVKNP	1050	
QY	829	TTIHKSP--	--DESTPELSA--EP--	TPKALENSPKPEGV-----	PTTKTPA	867
Db	1051	VKKWAPPWEDDEPSEPVSAPEBEKKTTPVLAKKAPTTPKATKPDSEAAADPVSGPTSKDKP	1110			
QY	868	AT-----	KPEMTTAKDKTTEHDRLTTPETT--	TAAPKWTKETATTITEKTTESKITATT	919	
Db	1111	LSKAPVEKPKPTTDPKDDKLSKPAKKKEKAPPAAPKPKWPVWDDDPDEADFATVPA	1170			
QY	920	TQVTSTTQDTPPKITTLTKTTLAPVTTTKTITITTEIMNKPEETAKPKDRATNSKAT	979			
Db	1171	PSKPEDTEDPADPLG-----	GPKTKDPK-----	LNKKAPEKPTPEK-----	1206	
QY	980	TPKPKQTAKPKKPTSTSKPKTMPRVKKPK-----	TTTPPKMWTSTM	1022		
Db	1207	-PKPEVSKPEKPTPEPPPK-AAAPKKWKPPWEDDPEADEADFTMPAKPKPDTEDP	1259			

RESULT 5

RESULT
T30826

T30826
nascen.

nascen
N. 21 + 2

N; Alte:

C:Spec:

C; Date

C; Access

R; Yotov

Genes I

Genes I
A:Title

A: Title
A: Before

A; Refer

A; Acces

A; Statu

A; Molec

A; Resic

and, respectively,

nascent polypeptide-associated complex alpha chain, muscle splice form gp220 - mouse

N;Alternate names: alpha-NAC protein
C:Species: Mus musculus (house mouse)
N;Alternate names: complex alpha chain, muscle splice form gp220 - mouse

C:Species: Mus musculus (house mouse)
C:Date: 22-Oct-1988

C:\Data\22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 24-Nov-1999

C; Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 24-Nov-1999
C; Accession: T30826

R; Yotov, W.V.; St-Arnaud, R.
Accession: 130826

R; IOTOV, W.V.; St-Arnaud, R.
Genes Dev. 10, 1763-1772, 1996

Genes Dev. 10, 1763-1772, 1996

A;Title: Differential splicing-in of a proline-rich exon converts alphaNAC into a muscivore-specific protein
A;Reference number: Z20889; MIM:06312450

A;Reference number: Z20889; MUID:96312450
A;Accession: T30826

A;Accession: T30826
A;Status: preliminary

A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA

A: Molecule type: DNA

```

molecule type: DNA
A;Residues: 1-2187 <YOT>

```

α, residues: 1-218 / <YOT>

J. Clin. Invest. 87, 77-82, 1991

A:Title: Human bronchus and intestine express the same mucin gene.
A:Reference number: A61257; MUID:91086481
A:Accession: A61257
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 'T',1925-1948,'TTS',1952-1954 <JAN>
A:Experimental source: bronchus
R:Xu, G.; Huan, L.; Khatra, I.; Sajjan, U.S.; McCool, D.; Jones, M.
Biochem. Biophys. Res. Commun. 183, 821-828, 1992
A:Title: Human intestinal mucin-like protein (MLP) is homologous with rat
A:Reference number: PQ0328; MUID:92198477
A:Accession: PQ0328
A:Molecule type: mRNA
A:Residues: 2328-2468 <XUG>
A:Cross-references: GB:M86523
A:Experimental source: small intestine
A:Accession: PQ0329
A:Molecule type: protein
A:Residues: 2328-2342,'K',2344-2354 <XUG1>
C:Genetics:
A:Gene: GDB:MUC2
A:Cross-references: GDB:120203; OMIM:158370
A:Map position: 11p15.5-11p15.5
C:Superfamily: von Willebrand factor; von Willebrand factor type A repeat
C:Keywords: glycoprotein; intestine; tandem repeat
F:2766-2834/Domain: von Willebrand factor type C repeat homology <WVC>

Query Match 14.7% Score 1073; DB 2; Length 3020;
Best Local Similarity 27.1% Pred. No. 4.3e-38;
Matches 374; Conservative 101; Mismatches 502; Indels 402; Gaps

QY 29 SCAGRCGGYSDATCNDYNCQHMEC-----CPDFKRVCTAAEVHNPTS 74
Db 1079 SCG--CDTG--GDCEGCSAVASYAQECTREGACVFWRTPDLCPF-----CDYNNPH 1128
QY 75 PSSKAKPPPGSGASQTIKSTT-----KRSPK--PPNKKTKKVISEB-- 115
Db 1129 ECEWHYPCGNRSPETCRTINGHSNISVSYLEGCVPRCKDRIYEDLKKCVTADRC 1188
QY 116 ---ITEH---SVSENESSSSSSSSSTIWKSSKNSAANRELKLVKDNKKNR 169
Db 1189 CYVEDTHYPGASVPTPECTCKVCVNSQVGRPEEG-----ILNQ 1231
QY 170 TKKKPTPKPPVVDAGSLDNGDF---KVTPDTSTQH---NKYSTPKITTAKPINR 223
Db 1232 TQ-----DGAFCVELCGNGVKEHFNCSITTRPSLTITFTTTITL 1273
QY 224 PSLPNSDTSKETSLTVNKTETVETKET--TTNK-----QTSDGKEKT 266
Db 1274 PTP-----TSFTTTTTTPTTSSTVLTSTPKLCLLSDWINEDHPSSGSDGDREP 1325
QY 267 TS---AKETQSIKTSARD----- 282
Db 1326 FDGVCGAPEDIECRSVKDPHLSLEHQGVQCDVSVGFICKNEQDFNGPGLCYDYKIR 1385
QY 283 -----LAPTSKVLAKPTPKAETTKGPAITTPKREPTTTPKPASTPKPEPTPT 332
Db 1386 VNCQWMDKCIITTPSPPTTTPSPPTTTLPTTTPSPPTTTTTPPTTTPPTTTPPTT 1445
QY 333 IKSAP--TTPKEAPTTTKSAPTTKPEAPTTTTPKPAETTKPEAPTTKPAETTKSAP 391
Db 1446 TTPLTPTTTPPTTSTTTTPPTTTPSP--PTTTPSPPTTTPSPPT--TTTTTTPPTTTPSP 1503
QY 392 -TTPKPEAPTTPKKAPTTTPKPEAPTTTPKPEPTTTPKPEAPTTTPKPAETTKPAETAPK 450
Db 1504 MTPITTPASTTTLPPTTTPSPPTTTTTPPTTTPPTTTPSPPTTTPPTTTPPTTTP 1562
QY 451 KPAPTTPKPEAPTTTPKPEAPTTTPKPEPTTTPKPEAPTTTKSAP--TTTKPEAPTTTKSAP- 508
Db 1563 SPPTTTPPTTTPPTTTPSPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTT 1622

Mon Apr 29 08:35:21 2002

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:31:52 ; Search time 114.61 Seconds
(without alignments)
907.900 Million cell updates/sec

Title: AA6
Perfect score: 7294
Sequence: 1 MANKTLPIYLLLSVFVQ.....ARATTSQTLKSVWVNC 1366

Scoring table:
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1073	14.7	3020	A43932	musin 2 precursor,
2	950	13.0	1664	T18262	S-layer protein -
3	863.5	11.8	1489	T31108	cyst germination s
4	846	11.6	1274	T16251	hypothetical prote
5	812	11.1	2187	T30826	nascent polypeptid
6	810.5	11.1	1367	S48478	glucan 1,4-alpha-g
7	784	10.7	1188	S49915	extensin-like prot
8	695	9.5	1151	T18535	high molecular mas
9	680	9.3	1229	T25697	hypothetical prote
10	672	9.2	1344	A35175	musin 1 precursor,
11	661.5	9.1	7962	T38346	elastin titin - hu
12	649	8.9	3570	T45025	musin MUC5B, trach
13	647.5	8.9	3507	T34513	hypothetical prote
14	643.5	8.8	990	T51618	nucleolar phosphop
15	633	8.7	489	T11622	extensin class 1 p
16	632	8.6	761	T23757	hypothetical prote
17	629	8.6	6642	T19431	protein UNC-89 - C
18	627.5	8.6	971	T75518	hypothetical prote
19	622.5	8.5	839	T29018	hypothetical prote
20	607.5	8.3	801	T27923	hypothetical prote
21	607.5	8.3	924	S27923	gene LF3 protein -
22	605	8.3	379	S50125	laryal glue protei
23	592.5	8.1	2232	T34434	hypothetical prote
24	568.5	7.8	1630	T35577	ascites siialoglyco
25	559.5	7.7	350	S22456	hydroxyproline-ric
26	559	7.7	875	S23760	polyphenolic adhes
27	556.5	7.6	856	T16543	hypothetical prote
28	551	7.6	620	S06733	hydroxyproline-ric
29	551	7.6	1087	QFMSH	neurofilament trip

30	545	7.5	873	2	A47283	calphotin - fruit
31	542	7.4	369	2	S20500	hydroxyproline-ric
32	538.5	7.4	416	2	JU0465	extensin precursor
33	537.5	7.4	1459	2	T32271	hypothetical prote
34	537	7.4	1072	1	A37421	neurofilament trip
35	532	7.3	756	2	T27642	hypothetical prote
36	530.5	7.3	813	2	S70795	vsaa protein precu
37	530.5	7.3	1162	2	JH0557	exo-alpha-sialidas
38	530	7.3	865	2	A47282	calcium-binding pr
39	526	7.2	3534	2	T42567	legument protein 2
40	522	7.2	328	2	JQ0985	hydroxyproline-ric
41	521.5	7.1	854	2	S02003	neurofilament trip
42	518	7.1	866	2	T45462	membrane glycoprot
43	518	7.1	1611	2	T38236	hypothetical prote
44	513	7.0	1832	2	T31113	musin-like glycopr
45	512.5	7.0	867	2	T45463	membrane glycoprot

ALIGNMENTS

RESULT 1
A43932
musin 2 precursor, intestinal - human (fragments)
N:Alternate names: musin SMUC-41
C:Species: Homo sapiens (man)
C:Date: 10-Mar-1993 #sequence_revision 12-Apr-1996 #text_change 05-Nov-1999
C:Accession: A49963; A45106; B45106; A43932; B3532; A61257; PQ0328; PQ0329
R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, B.; Kim, Y.S.
J. Biol. Chem. 269, 2440-2446, 1994
A:Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of t
A:Reference number: A49963; MUID:94132002
A:Accession: A49963
A:Molecule type: mRNA
A:Residues: 1-639 <GU1>
A:Cross-references: GB:L21998
R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Rothe, E.M.; Lagace, R.E.; Kim, Y.S.
J. Biol. Chem. 267, 21375-21383, 1992
A:Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both up
A:Reference number: A45106; MUID:93016075
A:Accession: A45106
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 626-1895 <GU2>
A:Cross-references: GB:M94131; NID:g186395; PIDN:AAA59163.1; PID:g186396
A:Note: sequence extracted from NCBI backbone (NCBIP:116706)
A:Accession: B45106
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 2037-3020 <GU3>
A:Cross-references: GB:M94132; NID:g186397; PIDN:AAA59164.1; PID:g186398
A:Experimental source: colon
A:Note: sequence extracted from NCBI backbone (NCBIP:116698)
R:Toribara, N.W.; Gum Jr., J.R.; Culhane, P.J.; Lagace, R.E.; Hicks, J.W.; Petersen,
J. Clin. Invest. 88, 1005-1013, 1991
A:Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polym
A:Reference number: A43932; MUID:91358717
A:Accession: A43932
A:Molecule type: DNA
A:Residues: 1343-1350, 'L', 1352-1411, 'S', 1413-1448, 'P', 1450-1503, 'T', 1505-1915 <TOR>
A:Cross-references: GB:M74027; NID:g188863; PIDN:AAA59875.1; PID:g188864
A:Note: sequence inconsistent with the nucleotide translation
A:Note: sequence extracted from NCBI backbone (NCBIP:55749, NCBIP:55750)
R:Gum, J.R.; Byrd, J.C.; Hicks, J.W.; Toribara, N.W.; Lamport, D.T.A.; Kim, Y.S.
J. Biol. Chem. 264, 6480-6487, 1989
A:Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and evi
A:Reference number: A33532; MUID:89197956
A:Accession: B33532
A:Molecule type: mRNA
A:Residues: 1916-2193 <GU4>
A:Cross-references: GB:M22405; NID:g188873; PIDN:AAA36334.1; PID:g188874
A:Experimental source: intestine
R:Janjy, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Kim, Y.S.; Basbaum, C.B.

Qy	837	TTTKEPTT-----IHKSPDESTPSELSPAETPKALENSKPEGVPYTTKTPAA-----	882
Db	703	DXRXPXPGSTAPXAHGVTSPDXRXPXPGSTAPXAHGVTSPDXRXPXPGSTAPXAHGVTSP	762
Qy	883	-TKPEMTTAKDK---TTERDLRTTPEMTTAAPKWKTKETATTETKTESKITATTTOVTS	938
Db	763	DXRXPXPGSTAPXAHGVTSPDXRXPXPGST--APXAHGVTSPDXRXPXPGSTAPXAHGVTSP	820
Qy	939	TTTQDTTTFMTTKTTLTTLAPKVTITTKTKITTEIMNKPPEETAKPKDRATNSKATTPKPK	998
Db	821	APDXRXPXG-----STAPXAHGVT-----SAPDXRXPXGSTAPXAHGVTSPDXRXPX	868
Qy	999	KPKAPKPKTSTKTKPKTPRVKPKTTPTRKWTSTMPELNPTSKIAEAMLOTTTRNQOT	1058
Db	869	-GSTAPXAHGVTSPDXRP--XPGSTAPXAHGVTSA--FDXRPX-----PGST	911
Qy	1059	-PNSKLVENVPKSEDAGGAGETPHMLLRPHVFMEVPTDMDYLPVPNOGIIINPLMSD	1117
Db	912	APXAHGVTSPDXRXPXPGSTAPXAH-----GVTSPDXRXPXGSTA---PXAHG	957
Qy	1118	ETWNICKGPKVDGLTT--LRNGTLVAPRGRHVFNWLSPEPPSPARITETWGIPIPID	1172
Db	958	VTSAPDXRXPXGSTAPXAHGVTSA-----PDXRPXPGSTAPXAHGVTSPDX	1003

Search completed: April 26, 2002, 16:16:52
Job time: 203 sec

OTHER INFORMATION: CCC, CCA, or CCG; and Ala = GCT, GCC, GCA, or GCG."

FEATURE: Peptide
NAME/KEY: Peptide
LOCATION: 144
OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa
OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC,
OTHER INFORMATION: or AGC; and Asn = AAT or AAC."

FEATURE: Peptide
NAME/KEY: Peptide
LOCATION: 147
OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC,
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."

FEATURE: Peptide
NAME/KEY: Peptide
LOCATION: 1, 21
OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
OTHER INFORMATION: 21 amino acid precursor sequence."
US-09-083-116-2

Query Match 5.7%; Score 424.5; DB 4; Length 2035;
Best Local Similarity 23.7%; Pred. No. 2.4e-18;
Matches 255; Conservative 69; Mismatches 511; Indels 241; Gaps 53;

QY 250 TSLAVNKETTVETKETTNNKQSTDCGKETSATQSIKTSKAD-LAPTSKVLAKPT 308
Db 16 TVLTV-----VTGSHASSTPGGKETSATQSSVPSSTKNAVMTSVLSSHS 65
QY 309 P-KAETTTKGP--ALTTPKEP-----TPTTPKEPASTTPKEPTTTIKSAP 351
Db 66 PGSSSTTGGQDVTLAPATEPASGAATGWQDVTVPVTRPALGSTTPPAHDVT---SAP 122
QY 352 TTPKEPAPTT-----KSAPTTTPKEP-----APTTPKEPAPTTTPKEPAPT 392
Db 123 ---DNKPAFGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVT 180
QY 393 TKE--PAPTTT-----KSAPTTTPKEPAPTTTPK-----KPAPTTTPKEPAPTTTPKEPTTT 439
Db 181 APDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVT 240
QY 440 PKE--PAPTTTPKEP-----APTTPKEPAPTTAPK-----KPAPTTTPKEPAPTTTPKEPAPT 486
Db 241 APDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVT 300
QY 487 TKEPSTTPKEPAPT--TKSAPTTTPKEPAPT-----TKSAPTTTPKEPSPPTTK-----E 535
Db 301 APDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVT 359
QY 536 PAPTTPKEPAPTTTPK-----KPAPTTTPKEPAPTTTPK-----EPAPTTPKPAKPAPKEPAPT 587
Db 360 SAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVT 419
QY 588 TPKEPAP-----TTPKKLTPTTPKEKLAPTTTPKEPAPTTPEELAPTTPEPTTPTPKEPAPT 643
Db 420 SAPDXRXPXGSTAPXAHGVT-----APDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVT 471
QY 644 TPKA-----AAPNTPKPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTPKGT--APTTLKEPAP 598
Db 472 APXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVT 530
QY 699 TTP-----KKPAKELAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTP 746
Db 531 TAPXAHGVTAPDXRXPXGSTAP--XAHGVTAPDXRXPXGSTAPXAHGVT-----APDXR 585
QY 747 KEPAPTTPK-----TAPTLKEPAPTT-----KKPAKELAPTTTPKGTSTTSD 792
Db 586 XPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAP--XAHGVTAPDX 644
QY 793 KPAP--TTPK-----ETAPTTTPKEPAPTTTP-----KKPAPTTTPETPTTSEVSTP 836
Db 645 RPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAP 702

Db 912 APXAHGVTAPDXRXPXGSTAPXAH-----GVTSAPDXRXPXGSTA-----PXAHG 957

QY 1118 ETNICKGKPVGLDT--LRNGTLVAFRGHYFWMLSPFSPSPARRITVWGIPIPID 1172
Db 958 VTISAPDXRXPXGSTAPXAHGVTSA-----PDXRXPXGSTAPXAHGVTISAPD 1003

RESULT 15

US-09-083-116-2
Sequence 2, Application US/09083116
Patent No. 6203795
GENERAL INFORMATION:
APPLICANT: CHAMON, Pierre
APPLICANT: KIENY, Marie-Paule
APPLICANT: LAHE, Richard
APPLICANT: HAREUVENI, Mara
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
TITLE OF INVENTION: TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/083,116
FILING DATE:
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/479,537
FILING DATE:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR91/00835
FILING DATE: 23-OCT-1991
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/039,320
FILING DATE: 04-APR-1993
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/403,576
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 017753-025
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2035 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Peptide
LOCATION: 128, 1899
OTHER INFORMATION: /note= "The amino acids spanning
OTHER INFORMATION: 128 to 1899 constitute a repeated region wherein the repeat
OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such
OTHER INFORMATION: repeats varies from 1 to 40."
FEATURE:
NAME/KEY: Peptide
LOCATION: 134
OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa
OTHER INFORMATION: Xaa Xaa which is the codon for Pro or Ala wherein Pro = CCT,

NUMBER OF SEQUENCES: 5
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
 STREET: P.O. Box 1404
 CITY: Alexandria
 STATE: Virginia
 COUNTRY: United States
 ZIP: 22313-1404
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/479,537A
 FILING DATE: 07-JUN-1995
 CLASSIFICATION: 514
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: FR 90/13101
 FILING DATE: 23-OCT-1990
 APPLICATION DATA:
 APPLICATION NUMBER: WO PCT/FR91/00835
 FILING DATE: 23-OCT-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/039,320
 FILING DATE: 04-APR-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/403,576
 FILING DATE: 14-MAR-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Teskin, Robin L.
 REGISTRATION NUMBER: 35,030
 REFERENCE/DOCKET NUMBER: 017753-025
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 836-6620
 TELEFAX: (703) 836-2021
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2035 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 FEATURE:
 NAME/KEY: Peptide
 LOCATION: 128..1899
 OTHER INFORMATION: /note= "The amino acids spanning 128 to 1899 constitute a repeated region wherein the repeat is 20 amino acids, 17 of which are fixed. The number of such repeats varies from 1 to 40."
 OTHER INFORMATION: repeats varies from 1 to 40."
 FEATURE:
 NAME/KEY: Peptide
 LOCATION: 134
 OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa
 OTHER INFORMATION: Xaa Xaa which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA, or CCG; and Ala = GCT, GCC, GCA, or GCG."
 OTHER INFORMATION: CCC, CCA, or CCG; and Ala = GCT, GCC, GCA, or GCG."
 FEATURE:
 NAME/KEY: Peptide
 LOCATION: 144
 OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa
 OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA, or ACG; and Asn = AAT or AAC."
 OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
 FEATURE:
 NAME/KEY: Peptide
 LOCATION: 147
 OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa
 OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA, or CCG; and Ala = GCT, GCC, GCA, or GCG."
 OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
 FEATURE:
 NAME/KEY: Peptide
 LOCATION: 1..21
 OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
 OTHER INFORMATION: 21 amino acid precursor sequence."

US-08-479-537A-2

Query Match 5.7%; Score 424.5; DB 2; Length 2035;
 Best Local Similarity 23.7%; Pred. No. 2.4e-18;
 Matches 255; Conservative 69; Mismatches 511; Indels 241; Gaps 53;

QY 250 TSLTVNKKETTVEKTTTNNKQTSDGKEKTSNAKETQSTIEKTSKD-LAPTSTKVLAAPT 308
 DB 16 TVLTV-----VTCGSHASSTPGCKETSATORSSVPSSTKNAVSMTSVLSHS 65
 QY 309 P-KAETTTKGP--ALTPKPEP-----TPTPKPEASTTPKEPTTTIKSAP 351
 DB 66 PGSGSSTQGDVTLAPATEPASGSAATWGQDVTSPVTRPALGSTTPPAHDVT---SAP 122
 QY 352 TTPKEPAPTTT-----KSAPTTPKEP-----APTTTKEPAPTTPKEPAPT 392
 DB 123 --DNKPAPGSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVT 180
 QY 393 TKE--PAPTTT-----KSAPTTPKEPAPTTPK-----KPAPTTPKEPAPTTPKEPTTT 439
 DB 181 APDXRXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVT 240
 QY 440 PKE--PAPTKEP-----APTTTKEPAPTAPK-----KPAPTTPKEPAPTTPKEPAPT 486
 DB 241 APDXRXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVT 300
 QY 487 TKEPSTTPKEPAPT--TKSAPTTPKEPAPT-----TKSAPTTPKEPAPTTPK-----E 535
 DB 301 APDXRP-XPGSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVT 359
 QY 536 PAPTTPKEPAPTTPK-----KPAPTTPKEPAPTTPK-----EPAPTTPKPAAPTAPKEAPT 587
 DB 360 SAPDXRXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVT 419
 QY 588 TPKETAP-----TPPKLTPPTPEKLAPTTPEKPAPTTPEELAPTTPEELAPTTPEEPAPT 643
 DB 420 SAPDXRXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVT 471
 QY 644 TPKA-----AAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKETAPTTPKET--APTTLKEPAP 698
 DB 472 APXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGSTAPX 530
 QY 699 TTP-----KKPAPKELAPTTTKEPTSTSDKPAP---TTPKGTAPTTPKEPAPTTP 746
 DB 531 TAPXAHGVTSPADXPXPGSTAP-XAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGSTAP 585
 QY 747 KEPAPTTPKG-----TAPTTLKEPAPTTP-----KKPAPKELAPTTTKEPTSTSD 792
 DB 586 PXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGSTAP-XAHGVTSPADXP 644
 QY 793 KPAP--TTPK-----ETAPTTPKEPAPTTP-----KKPAPTTPPETPTTSEYSTP 836
 DB 645 RPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPADXP 702
 QY 837 TTTKEPTT-----IHKSPDESTPELSAETPKALENSPKPGVPTTKTPAA----- 802
 DB 703 DXRPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPAP 762
 QY 883 -TKPEMTTAKDK---TTERDLRTPTTTAAPKMKETATTTETTESKITATTTQVTS 938
 DB 763 DXRPXPGSTAPXAHGVTSPADXPXPGST--APXAHGVTSPADXPXPGSTAPXAHGVT 820
 QY 939 TTTQDTPPFKITTLTKTTTLPAPKVTTKKTLTTTTEIMNKPEETAKPKDRATNSKATPKPQ 998
 DB 821 APDXRXPXG-----STAPXAHGVT-----SAPDXRXPXPGSTAPXAHGVTSPADXPX 868
 QY 999 KPTKAPKPTSTKKPTMPRVKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTTTPNOT 1058
 DB 869 -GSTAPXAHGVTSPADXPX--XPGSTAPXAHGVTSA-PDXRPX-----PGST 911
 QY 1059 -PNSKLVENVNPKSEDAGAGETPHMLLRPHVFMPEVTPMDVLPVPRVNOGIINPMLSD 1117

APPLICATION NUMBER: 08/479,537
FILING DATE: 23-OCT-1991
PRIORITY DATE: 04-APR-1993
PRIORITY APPLICATION NUMBER: US 08/039,320
PRIORITY APPLICATION DATE: 04-APR-1993
PRIORITY APPLICATION NUMBER: US 08/403,576
PRIORITY APPLICATION DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 017753-025
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1867 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Peptide
LOCATION: 128..1727
OTHER INFORMATION: /note= "The amino acids spanning 128 to 1727 constitute a repeated region wherein the repeat is 20 amino acids, 17 of which are fixed. The number of such repeats varies from 1 to 40."
OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 134
OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa
OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA or ACG; and Asn = AAT or AAC."
FEATURE:
NAME/KEY: Peptide
LOCATION: 147
OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 147
OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
OTHER INFORMATION: 21 amino acid precursor sequence."
US-09-083-116-5
Query Match 5.7%; Score 424.5; DB 4; Length 1867;
Best Local Similarity 23.7%; Pred. No. 2.1e-18;
Matches 255; Conservative 69; Mismatches 511; Indels 241; Gaps 53;
QY 250 TSLVTKETTTTETKTTTNGKTTTSDGKTTTSAKETQSIKTSKAD-LAPTSKVLAKPT 308
DB 16 TVLTV-----VTGSGHASSTPGGEKETSATQSSVPSSTKNAVMTSSVLSSHS 65
QY 309 P-KAETTTKGP--ALTTPKEP-----TPTTPKEPASTTPKEPTTIKAP 351
DB 66 PGSGSSTTQGVDTLAPATAPASGSAATWGQDVTSVPTVRPALGSTTPPAHDVT---SAP 122
QY 352 TTPKEPAPTT-----KSAPTTKEP-----APTTPKEPAPTTTPKEPAPT 392
DB 123 --DNKPAPGSTAPXAHGVTAPDXRXPFGSTAPXAHGVTAPDXRXPFGSTAPXAHGVT 180

QY 393 TKE--PAPTTT-----KSAPTTTPKEPAPTTPK-----KPAPTTTPKEPAPTTTPKEPTTT 439
DB 181 APDXRXPFGSTAPXAHGVTAPDXRXPFGSTAPXAHGVTAPDXRXPFGSTAPXAHGVT 240
QY 440 PKE--PAPTTKEP-----APTTPKEPAPTPK-----KPAPTTTPKEPAPTTTPKEPAPT 486
DB 241 APDXRXPFGSTAPXAHGVTAPDXRXPFGSTAPXAHGVTAPDXRXPFGSTAPXAHGVT 300
QY 487 TKEPSPTTPKEPAPTT--TKSAPTTTPKEPAPTT-----TKSAPTTTPKEPSPTTTK-----E 535
DB 301 APDXRP--XPGSTAPXAHGVTAPDXRXPFGSTAPXAHGVTAPDXRXPFGSTAPXAHGVT 359
QY 536 PAPTTTPKEPAPTTPK-----KPAPTTTPKEPAPTTPK-----EPAPTTTPKAPAPKAPAPT 587
DB 360 SAPDXRXPFGSTAPXAHGVTAPDXRXPFGSTAPXAHGVTAPDXRXPFGSTAPXAHGVT 419
QY 588 TPKETAP-----TTPKLTPTTPKEKAPTTTPKEKAPTTTPKEKAPTTTPKEKAPTTTPKEKAPT 643
DB 420 SAPDXRXPFGSTAPXAHGVT-----APDXRXPFGSTAPXAHGVTAPDXRXPFGSTAPXAHGVT 471
QY 644 TPKA-----AAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPT 698
DB 472 APXAHGVTAPDXRXPFGSTAPXAHGVTAPDXRXPFGSTAPXAHGVTAPDXRXPFGSTAPXAHGVT 530
QY 699 TTP-----KKPAPKELAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPT 746
DB 531 TAPXAHGVTAPDXRXPFGSTAP--XAHGVTAPDXRXPFGSTAPXAHGVT-----APDXR 585
QY 747 KEPAPTTPKG-----TAPTTKEPAPTT-----KKPAPKELAPTTTPKEPAPTTTPKEPAPT 792
DB 586 XPXGSTAPXAHGVTAPDXRXPFGSTAPXAHGVTAPDXRXPFGSTAPXAHGVTAPDXRXPFGSTAPXAHGVT 644
QY 793 KPAP--TTPK-----ETAPTTTPKEPAPTT-----KKPAPTTTPPTTPPTTPPTTPPTTP 836
DB 645 RXPFGSTAPXAHGVTAPDXRXPFGSTAPXAHGVTAPDXRXPFGSTAPXAHGVTAPDXRXPFGSTAPXAHGVT 702
QY 837 TTTTKEPT-----IHKSPDESTELSAEPTPKALENSPKPEPGVPTTKTPAA----- 882
DB 703 DXRXPFGSTAPXAHGVTAPDXRXPFGSTAPXAHGVTAPDXRXPFGSTAPXAHGVTAPDXRXPFGSTAPXAHGVT 762
QY 883 --TKPEMTTAKDK---TTERDLRTTTPETTAAPKMTKTATTTTEKTESKITATTQTTS 938
DB 763 DXRXPFGSTAPXAHGVTAPDXRXPFGST--APXAHGVTAPDXRXPFGSTAPXAHGVT 820
QY 939 TTTQDTPPEKITTLKTTTLAPXVTTTKTTTTEIMNPEETAKPKDRATNSKATTPKPQ 998
DB 821 APDXRXPFG-----STAPXAHGVT-----SAPDXRXPFGSTAPXAHGVTAPDXRXPFGSTAPXAHGVT 868
QY 999 KPTKAPKPTSTKPKTMRVRKPKTTTPPKMTSTMPKLTSTRIAEAMLOTTTRPNOT 1058
DB 869 -GSTAPXAHGVTAPDXRXP--XPGSTAPXAHGVTSA-PDXRXP-----PGST 911
QY 1059 -PNSKLVEVNPKSEDAAGGETPHMLLRPHVPMPEVTPDMYLPVPMNOGIIINPLMSD 1117
DB 912 APXAHGVTAPDXRXPFGSTAPXAH-----GVTAPDXRXPFGSTA-----PXAHG 957
QY 1118 ETNICKPKPVDGLTT-LRNGTLVAERGHYFWMLSPSPARRITEVVGIFSPID 1172
DB 958 VTSAPDXRXPFGSTAPXAHGVTSA-----PDXRXPFGSTAPXAHGVTAPDXRXPFGSTAPXAHGVT 1003
RESULT 14
US-08-479-537A-2
; Sequence 2, Application US/08479537A
; Patent No. 5861381
; GENERAL INFORMATION:
; APPLICANT: CHAMON, Pierre
; APPLICANT: KIENY, Marie-Paule
; APPLICANT: LATHE, Richard
; APPLICANT: HAREUVENI, Mara
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
; TREATMENT OR PREVENTION OF A MALIGNANT TUMOR

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 1867 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

FEATURE:

NAME/KEY: Peptide

LOCATION: 128..1727

OTHER INFORMATION: /note= "The amino acids spanning

128 to 1727 constitute a repeated region wherein the repeat

of 20 amino acids, 17 of which are fixed. The number of such

FEATURE: repeats varies from 1 to 40."

NAME/KEY: Peptide

LOCATION: 134

OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa

which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA

or CCG; and Ala = GCT, GCC, GCA, or GCG."

FEATURE:

NAME/KEY: Peptide

LOCATION: 144

OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa

which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA

or ACG; and Asn = AAT or AAC."

FEATURE:

NAME/KEY: Peptide

LOCATION: 147

OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa

which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA

or CCG; and Ala = GCT, GCC, GCA, or GCG."

FEATURE:

NAME/KEY: Peptide

LOCATION: 1..21

OTHER INFORMATION: /note= "Amino acids 1 to 21 are a

OTHER INFORMATION: 21 amino acid precursor sequence."

US-08-479-537A-5

Query Match

Best Local Similarity 5.7%; Score 424.5; DB 2; Length 1867;

Matches 255; Conservative 69; Mismatches 511; Indels 241; Gaps 53;

```

QY 250 TSLTVNKTETVETKETTNNKSTDGKETSATQSIKTSKAD-LAPTSKVLAKPT 308
DB 16 TVLTV-----VTGSGHASTPGGKETSATQSVSPSSSEKNAVMTSSVLSHS 65
QY 309 P-KAETTTKGP-ALTTKPEP-----TPTTPKPASTTTKPEPTTIKSP 351
DB 66 PGSGSTTGQDVTLPATEPAGSAAATWGQDVTSPVTRPALGSTTTPAHDVT---SAP 122
QY 352 TTPKPEAPTTT-----KSAPTTTKEP-----APTTTKEPAPTTKPEAPTT 392
DB 123 --DNKPAAGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVT 180
QY 393 TKE--PAPTTT-----KSAPTTTKEPAPTTK-----KPAPTTTKEPAPTTKPEPTT 439
DB 181 APDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVT 240
QY 440 PRE--PAPTTKEP-----APTTTKEPAPTAPE-----KPAPTTTKEPAPTTKPEAPTT 486
DB 241 APDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVT 300
QY 487 TKEPSTTPKPEAPTT--TKSAPTTTKEPAPTT-----TKSAPTTTKEPSPPTTK-----E 535
DB 301 APDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVT 359
QY 536 PAPTTTKEPAPTTK-----KPAPTTTKEPAPTTK-----EPAPTTTKEPAPTTKPEAPTT 587
DB 360 SAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVT 419
QY 588 TPKEAP-----TTPKKLTPTTPEKLAPTTPEKAPTTTPEETPTTPEEPAPT 643

```

```

DB 420 SAPDXRXPXGSTAPXAHGVT-----APDXRXPXGSTAPXAHGVTAPDXRXP-----PGST 471
QY 644 TPKA-----AAPNTPKPEAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 698
DB 472 APXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPX 530
QY 699 TTP-----KKPAPKELAPTTTKEPTSTTSDKAP--TTPKGTAPTTTKEPAPTTT 746
DB 531 TAPXAHGVTAPDXRXPXGSTAP--XAHGVTAPDXRXPXGSTAPXAHGVT-----APDXR 585
QY 747 KEPAPTTPKG---TAPTTLKPEAPTT-----KKPAPKELAPTTTKEPTSTTSD 792
DB 586 PXPGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAP--XAHGVTAPDX 644
QY 793 KPAP--TTPK-----ETAPTTTKEPAPTT-----KKPAPTTTKEPTSTTSEVSTP 836
DB 645 RXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXP 702
QY 837 TTTTKEPT-----IHKSPDESTPELSNEPTPKALENSPKPEGVPVTTKTPAA----- 882
DB 703 DXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAP 762
QY 883 --TKPEMTTAKDK---TTERDLRTTPTTAAPKMTKETATTTEKTTESKITATTTOVTS 938
DB 763 DXRXPXGSTAPXAHGVTAPDXRXPXGST--APXAHGVTAPDXRXPXGSTAPXAHGVT 820
QY 939 TTTQDTTPEKITLTKTTTLPKVTITTTTITTTIMNKPEETAKPKDRATNSKATTPKPO 998
DB 821 APDXRXPXG-----STAPXAHGVT-----SAPDXRXPXGSTAPXAHGVTAPDXRXP- 868
QY 999 KPTKAPKPTSTKKPKTMPRVKPKTTPPKMTSTMPKLNPTSRIAEAMLTITRPNQT 1058
DB 869 -GSTAPXAHGVTAPDXRXP--XPGSTAPXAHGVTSA-PDXRXP-----PGST 911
QY 1059 -PNSKLVENVPKSEDAGAGETPHMLLRPHVFMPEVTPMDYLPVNPNGIINPMLSD 1117
DB 912 APXAHGVTAPDXRXPXGSTAPXAH-----GVTAPDXRXPXGSTA---PXAHG 957
QY 1118 ETNICKGKPVGLTT-LRNGTLVAFRGHYFWMLSPPSPSPARRITEVWGIPSPID 1172
DB 958 VTSAPDXRXPXGSTAPXAHGVTSA-----PDXRXPXGSTAPXAHGVTAPDXRXP 1003

```

RESULT 13

US-09-083-116-5

; Sequence 5, Application US/09083116

; Patent No. 6203795

; GENERAL INFORMATION:

; APPLICANT: CHAMON, Pierre

; APPLICANT: KIENY, Marie-Paule

; APPLICANT: LATHE, Richard

; APPLICANT: HAREUVENT, Mara

; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE

; TITLE OF INVENTION: TREATMENT OR PREVENTION OF A MALIGNANT TUMOR

; NUMBER OF SEQUENCES: 5

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.

; STREET: P.O. Box 1404

; CITY: Alexandria

; STATE: Virginia

; COUNTRY: United States

; ZIP: 22313-1404

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/083.116

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

[illegible]

Db 678 ---APATTPEDDDIDP---PLPNDPINP 699

RESULT 6
US-08-928-361B-6
; Sequence 6, Application US/08928361B
; Patent No. 6071518
; GENERAL INFORMATION:
; APPLICANT: Petersen, Carolyn
; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
; THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
; TITLE OF INVENTION: SPECIES INFECTIONS
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PETERS, VERNY, JONES & BIKSA
; STREET: 385 Sherman Avenue, Suite 6
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-1840
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928,361B
; FILING DATE: 12-SEP-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA: US 60/026,062
; APPLICATION NUMBER:
; FILING DATE: 13-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: VERNY, Hana
; REGISTRATION NUMBER: 30,518
; REFERENCE/DOCKET NUMBER: 480.76-1(HV)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-324-1677
; TELEFAX: 650-324-1678
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1721 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-928-361B-6

Query Match 6.7%; Score 498.5; DB 4; Length 786;
Best Local Similarity 29.1%; Pred. No. 2e-23;
Matches 200; Conservative 25; Mismatches 207; Indels 255; Gaps 32;

307 PTPKAPATTGKALTPKPTTPKPEASTTPKKEFTTTIKSAPTTPKEPAPTTKSP 366
Db 93 PGPAB-TTQAPAT---QATTQAPTTT-----QAPTTTQ-ATTTQAP 135
367 TTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 426
Db 136 TTTQ--APTQ--ATT---QAPTTTQ-APTTPQAPTTPQ-APTQ--APTQ-- 183
QY 427 PATTKEPTTPKPEAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT- 483
Db 184 -APTTPQAPTTPQAPTTPAAPTTPAAPTTPAAPTTPAAPTTPAAPTTPAAPT 241
QY 484 -----PTTKEPSTTPK- 496
Db 242 ELLPNGCPADFIIHLIPHDKYCNLFYQCSNGYTFEQRCEPGLYFNYPVORCD 301
QY 497 -----EPATTT----- 503
Db 302 DGEISPAPVTEGNEDEDIDIGLLDNGCPANFEIDMLPHGNRCDDKYQCVHGN 361
QY 504 -----KSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 536
Db 362 CGAGTFHFELQOCDHTELVCGLPGGESEVDVDEDA-CTGWYCPEPIEWEPLNG 420
QY 537 APTT-----PKEP-----APTTPKAPTT----- 556
Db 421 ADFSIDHLLPHESDCGOYLQCVHGOITARPCCGNLHFSPTATQSCSPVITAGCQ 480
QY 557 ---PKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 613
Db 481 NQCTSTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 533
QY 614 EKPAPTTPEELAPTTPEPTTPEEPAPTTPEEPAPTTPEEPAPTTPEEPAPTT 671
Db 534 --PAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPT 587
QY 672 PAPTTPKETATPTTKGTAPTTLKEPAPTTTPKKAPKELAPTTTKEPTSTSDK 731
Db 588 AAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 632
QY 732 GPAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 791
Db 633 TAAPT--AAPAPT--VVPPTVVPPTVVPPTVVPPTVVPPTVVPPTVVPPT 677
QY 792 DKPAPTTPKETATPTTKGTAPTTLKEPAPTTTPKKAPKELAPTTTKEPT 818

Query Match 6.7%; Score 498.5; DB 3; Length 1721;
Best Local Similarity 21.8%; Pred. No. 5e-23;
Matches 300; Conservative 94; Mismatches 510; Indels 469; Gaps 44;

QY 107 STTKRSPKPPNKKYKVKVIESEIIEHSVSENQESSSSSSSSSSSSSSSSSS 160
Db 38 SITSGELKDPNKQAT-----ISGRSCGKQGYSIDSSGFRVDSITGLPT 83
QY 161 --SKNSAANRELQKKLVKXDNKKNRTKPKPPVVDVDEAGSLDNGDPKVTTP 218
Db 84 DPYSNCPFN-----PVTGNLVSRTGKTIPN---TVAGVYRSN-ETKTTPSANT 129
QY 219 HNKVSTSPKITTAKP-----INPRSLPPNSDTSKE-----TSLTVN 255
Db 130 YAGVYRSNETKTTPSANTNLLVDPKINAPCNSENSEFQGGIFDMGSKVYIPYTK 189
QY 256 KEITVETKETTTTNTKOTSTDGKEKTTSAKETQSIKTSARDLAPTSKVLAK 315
Db 190 KHTTT 211
QY 316 KGPALATPKPEPTTTPKEPASTTTPKEPTTTIKSAPTTPKEPAPTTTKEP 375
Db 212 TT 271

```
; Sequence 5, Application US/08700651B
; Patent No. 6015882
; GENERAL INFORMATION:
; APPLICANT: PETERSEN, CAROLYN
; APPLICANT: LEECH, JAMES
; APPLICANT: NELSON, RICHARD, C.
; APPLICANT: GUT, JIRI
; TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS
; TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum
; FILE REFERENCE: 480.19-4(HV)
; CURRENT APPLICATION NUMBER: US/08700.651B
; CURRENT FILING DATE: 1997-08-14
; EARLIER APPLICATION NUMBER: 08/415.751
; EARLIER FILING DATE: 1995-04-03
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 1721
; TYPE: PRT
; ORGANISM: Cryptosporidium parvum
; US-08-700-651-5

Query Match          6.7%; Score 499.5; DB 3; Length 1721;
Best Local Similarity 21.8%; Pred. No. 4.4e-23;
Matches 300; Conservative 94; Mismatches 510; Indels 469; Gaps 44;

QY 107 STTKRSPKPKKTKKVIKIEEITEEHSVSENOESSSSSSSSSTIWKIIS-----160
DB 38 SITSGERKDPKQAT-----ISGSRSCGWKOGYSIDSSTGFRVDSITGLPT 83
QY 161 --SKNSAARLEQLKLVKDKKNTKPKPPVVDVDEAGSLDNGDKFKVTTPTDSTTQ 218
DB 84 DYSNCFPN-----PVTGNLVSRSKSTGKIPN---TYAGVYRSN-ETKTEPSANT--129
QY 219 HNKYSTSPKTIKAP-----INRPSLPNSDTSKE-----TSLTVN 255
DB 130 YAGVYRSNETKTEPSANTNELLVDPKINAPCNSENSFEGQIFDMGSKYVYIPYTKCVG 189
QY 256 KETTVEIKETTNKQSTGCKEKTSAKETQSIKTSKDLAPTSKVLAKPPEKATTT 315
DB 190 KHTTTTTTTTTTTTTTTT-----TST 211
QY 316 KGPALTTPKEPTPTTPKEPASTTPKEPTPTIKSAPTTPKEPATTTKSAPTTPKEPAPT 375
DB 212 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 271
QY 376 TKKEPAPTTPKEPAPTTPKEPAPTTPKSPAPTTPKEPAPTTPKKEPAPTTPKEPAPTTPKEP 435
DB 272 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 331
QY 436 TPPTTPKEPAPTTPKEPAPTTPKSPAPTTPKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPT 495
DB 332 TTTTTPKPTTTTTTTTTT-----TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 377
QY 496 KEPAPTTTSAPTTPKEPAPTTPKSPAPTTPKE-----527
DB 378 TKPTTTTTTTTTTTTTTKPTTTTATTTTTTSETSVIKPDEWCWLEKNGEAKGATVYG 437
QY 528 -----PSPTT-----TKE 535
DB 438 VIKGOGRIENGMAFTWIPNDTHVFRFKVKVDGNTISVRCRKGAGKLEFPDRSLDFIP 497
QY 536 P-----APTTPKEPAPTTPKKEPAPTTPKE-----559
DB 498 PVAGHNSCSIIIVGSDGKIHVSPYSGKDVSLIS----APIQSELNFNEVCDCTAKYG 553
QY 560 ----PAPTTPKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPK 616
DB 554 AIHSGYQTSADVTTTAKPTTT--TGAPGQPTTTTSGPSKPTTTTAKTTTT--607
QY 617 ABPTPEELAPTTPKEPTTPPEEAPTTPKAAAPTTPKEPAPTTPKKEPAPTTPKKEPAPTTP 676
```

```
DB 608 --TTLNPIITTTOKPPTTT-----TT---KVPKPPATTTTTLKPIVTTTTTKATTT 656
QY 677 PKETAPPTPKGTAPTTLKEPAPTTPKPKAPK-----ELAPTTTKEPTS-----719
DB 657 TTTTVPPT-----TTTTKRDEMTTTTTPLDIGDIEITPIEKMLDKYTRMYDVNSGL 711
QY 720 --TTSKDPAPTTPKG-TAPITPKEPAPTTPKEPAPTTPKGTAPTTLKE---PAPTTPKPK 773
DB 712 LLDNSDEPIPGSQAGQIADTSLNLPVQTHKSTGLPDPVMVGLPDPKSGNLVHPYNTQM 771
QY 774 APKELAPTTPKGTSTTSKDPAPTTPKETAPTTPKKEPAPTTPKKEPAPTTPPEPTTSEV 833
DB 772 SGLSVSYLAALKNLTVDTDEYG--LPIDTLTGYPDPVSLIPN--PETGEUFDPISDSI 827
QY 834 STPT-----TTKEPTTIHKS-----PDEST- 853
DB 828 MNGTIAGIVSGISASESLLSKSALIDPATNMVVGEGLLNPATGVMIPGFLGSEQTO 887
QY 854 -----PELSAEPTRALENSPREPGVPTTKTPAATKPKEMTTAKOKTTTER---898
DB 888 FSPEDGGIIPPEVAANADKELSLP--PSVP-----ESIP-----KDKKIDISIE 934
QY 899 -----DLRTPPTTTAAPKMTKETATTTTEKTTESKITATT-TQVSTTTQD 943
DB 935 LMVDIESGRLIGQVSKRRPIPGSIAGDLNIPMKTPTQDTSVTKGPDPTTGLPFPNPPTGHL 994
QY 944 TTPFKITITLTKTTLAPKVTITTKKTTITTEIMNKP--EETAKPKO-----RATNSKATT 994
DB 995 INPTNNNTWDSFAGAYKAVSNGIKTDNVYGLVGEITGLPKDPSIPENSTTGELVD 1054
QY 995 PKPQPTKAPKPTSTKPKTMP-----RVKPKPTTTPPKMTS 1033
DB 1055 PSTGKPINNSTAGIVSGKPLPIEDENGLNLPDPSTNLPIDGNNOLVNPENSTVSGSTS 1114
QY 1034 --TMPE-----LNPTSRIAEAMLQTTTRPNQTPNSKL 1063
DB 1115 GTTKPKPGIVNGGVDPDEAKQADKGLIVPPTNSINKDPVTNTQVSTNTGNI--1172
QY 1064 VEVNPKSEDAGAEGETPHMLLRPHVEMPEVTPD-----MDYLRVNPQGIINP--1113
DB 1173 --INP--ETGKVIKPSLPGSLNYPSENFPOQTDEITGKPDVTNGLPYDSTGIIIDPAT 1228
QY 1114 -----MLSDETNICNGK-----PVDGLTTLRNGTLV 1139
DB 1229 KLPIPGSVAGDEILTEVLNITTTDEVTLGTLPLETGLPRDPVSGLPQLPNGTLV 1281
```

RESULT 5

```
US-09-103-429A-3
; Sequence 3, Application US/09103429A
; Patent No. 6187558
; GENERAL INFORMATION:
; APPLICANT: Granados, Robert R
; APPLICANT: Wang, Ping
; TITLE OF INVENTION: A No. 6187558el Invertebrate Intestinal Mucin
; TITLE OF INVENTION: cDNA and Related Products and Methods
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Pinnisi & Michaels, P.C.
; STREET: 118 No. 6187558th Tioga
; CITY: Ithaca
; STATE: NY
; COUNTRY: USA
; ZIP: 14850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/103.429A
; FILING DATE: 24-JUN-1998
```


Matches	226;	Conservative	33;	Mismatches	211;	Indels	342;	Gaps	40;
QY	300	TSKYLAKPTKAEATTTKGGALATTPKPEPTTTTKEPASITPKPEPTTIKSAP	-----TTP	354					
Db	9	TALGLVAAREVDSAEKNPALHEPHDXP	-PAEQXLLLPXEYDCTKYCYEGLKFIAP	66					
QY	355	KEAPATT-----TKSAPTTKPEARTTTTKEPAPTTTPKPEAPTTTK	402						
Db	67	RDCAPGTEKFSQATCVHAALAGCTLGGPAETT	---QAPATTQ--APTITQ--APTITQ--APTITQ	119					
QY	403	SAPTTKPEAPPTPKKPAATTTPKPEAPTTPKPEPTTTPKPEAPTTTPKPEAPTA	462						
Db	120	QAPTITTOATTT-----QAPTITQAPTIT-----QAPTITQAPTIT	-----156						
QY	463	PKKPAATTTPKPEAPTTTPKPEAPTTTPKPEAPTTTPKPEAPTTTPKPEAPTTTKSAP	522						
Db	157	---QAPTITQ--APTITQ--APTITQ-----APTITQ--APTITQ--APTITQ	198						
QY	523	TTPKESPTTTKEAPPTP--KEPAPTTPKKPAATTTPKPEAPTTTPK	-----566						
Db	199	TTPAATTAAAT--TPAATTAAATTPAATTGVPAPTS	-----APVPPICELLPNGCPADFD	253					
QY	567	-----	-----EPAPTTTK	574					
Db	254	IHLHPDKYCNLFQCSNGYTFEQRCEGILYFNYPVORCDSPANVECDGEISPA	PVTE	313					
QY	575	K-----PA-----	-----577						
Db	314	GNEDEDIDIGDLLDNGCPANFEIDWLLPHGNCRDKYYQCVHGNLVERRCCAGTHFSE	LQ	373					
QY	578	-----PTAP--KEPAPT-----	TPKET	592					
Db	374	QCDHIELVGCTLPGGESEVDVEDACTGWCTPEIEWEPLNGCPADFSIDLHPHES	433						
QY	593	-----APTPKKLPTT-----	-----PEKLAPTTPEKPA	618					
Db	434	DGGVQLQCVHGQTIARPCPGNLHFSPATQCESPVTAGCQVFECDSDNOCTSTA	ATAAP	493					
QY	619	TTPBELAPTTPEEPTTPTEEPAPTTTAKAANTTPKPEAPTTTPKPEAPTTTPK	678						
Db	494	TAAPTAAPTAAPTAAPTSTVPPPA--TPATAAPVPTTAIPT-----PAPTAAPTAA	548						
QY	679	ETAPTT---PKGTAPTTLKEPAPTTPKKPAKELAPTTTKEPTTSTSDKPA	735						
Db	549	PESTTVVPTTAAPTA-----APT-----AVPEIPTVSAPTAAPT--	AAPTAAPTAAP	598					
QY	736	TTPKPEAPTTTPKPEAPTTPKGTAPTTLKEPAPTTPKKPAKELAPTTTKGPT	ST-TSDKP	794					
Db	599	TTAVPEIPTT-----VTSPTTAAPT--AAPAPNT-----	TVVPTTAAPT	641					
QY	795	APTPPKETAPTTPKPEAPTTTPKPEAPTTTETPPPTTSEVSTTTPKPTTIHKS	DESTEP	854					
Db	642	AP-----NTVTAPPTAAPT--AAPAPNTTVVPTTAAPTAAAP-----	PTVAH-----AP	685					
QY	855	ELSAPTPKALENSPKPGVPTTKTPAATKPE	886						
Db	686	NTTAAP-----VTTTSAPATT--PE	703						

RESULT 2
US-08-928-361B-5
: Sequence 5, Application US/08928361B
: Patent No. 6071518
: GENERAL INFORMATION:
: APPLICANT: Petersen, Carolyn
: TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
: TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
: TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
: TITLE OF INVENTION: SPECIES INFECTIONS
: NUMBER OF SEQUENCES: 30
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: PETERS, VERNY, JONES & BIKSA
: STREET: 385 Sherman Avenue, Suite 6

CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306-1840
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/928,361B
FILING DATE: 12-SEP-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/026,062
FILING DATE: 13-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Vermy, Hans
REGISTRATION NUMBER: 30,518
REFERENCE/DOCKET NUMBER: 480.76-1(HV)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-324-1677
TELEFAX: 650-324-1678
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1837 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-928-361B-5

[illegible]

Mon Apr 29 08:35:42 2002

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model
Run on: April 26, 2002, 16:16:35 ; Search time 49.78 Seconds
(without alignments)
623.836 Million cell updates/sec

Title: us-09-556-246-1-copy_25_1404
Perfect score: 7410
Sequence: 1 QDLSSCAGRCGEGYSRATC.....ARAITRSQTLKVKVYNCP 1380

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PCTUS.COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	522	7.0	805	4	US-09-103-429A-4
2	513	6.9	1837	3	US-08-928-361B-5
3	506.5	6.8	744	6	5202236-25
4	499.5	6.7	1721	3	US-08-700-651-5
5	498.5	6.7	1721	3	US-09-103-429A-3
6	498.5	6.7	1721	3	US-08-928-361B-6
7	488.5	6.6	826	1	US-07-638-431-2
8	488.5	6.6	826	1	PCT-US92-00018-2
9	476.5	6.4	652	6	5202236-13
10	452	6.1	960	4	US-09-219-849-5
11	425.5	5.7	1185	4	US-09-041-886-23
12	424.5	5.7	1867	2	US-08-479-537A-5
13	424.5	5.7	1867	4	US-09-083-116-5
14	424.5	5.7	2035	2	US-08-479-537A-2
15	424.5	5.7	2035	4	US-09-083-116-2
16	419.5	5.7	2476	2	US-08-276-967-2
17	417	5.6	829	1	US-08-642-255-132
18	417	5.6	829	1	US-08-397-633A-53
19	417	5.6	837	1	US-08-175-155-68
20	417	5.6	837	1	US-08-477-509B-103
21	417	5.6	837	1	US-08-642-255-101
22	417	5.6	837	2	US-08-707-237A-75
23	417	5.6	837	3	US-08-482-085B-103
24	417	5.6	897	1	US-08-397-633A-50
25	413.5	5.6	907	3	PCT-US95-04611A-19
26	413.5	5.6	907	5	PCT-US95-04611A-19
27	404.5	5.5	408	1	US-07-609-716-65

28	404.5	5.5	408	4	US-08-475-411A-65	Sequence 65, Appl
29	404.5	5.5	408	4	US-08-478-029A-65	Sequence 65, Appl
30	398.5	5.4	682	1	US-08-642-255-126	Sequence 126, App
31	398.5	5.4	682	1	US-08-397-633A-36	Sequence 36, Appl
32	384	5.2	1537	3	US-08-325-267A-2	Sequence 2, Appl
33	383.5	5.2	1231	3	US-08-904-263A-4	Sequence 4, Appl
34	381	5.1	1848	4	US-08-296-791-6	Sequence 6, Appl
35	381	5.1	1848	5	PCT-US95-10661A-6	Sequence 6, Appl
36	376	5.1	960	4	US-09-219-849-6	Sequence 6, Appl
37	373	5.0	1786	4	US-08-973-462-8	Sequence 84, Appl
38	371	5.0	761	2	US-08-707-237A-84	Sequence 114, App
39	371	5.0	762	1	US-08-642-255-114	Sequence 26, Appl
40	371	5.0	762	1	US-08-397-633A-26	Sequence 62, Appl
41	370.5	5.0	1064	1	US-08-642-255-62	Sequence 28, Appl
42	370.5	5.0	1187	3	US-08-320-559-28	Sequence 28, Appl
43	370.5	5.0	1187	5	US-08-545-860D-28	Sequence 28, Appl
44	370.5	5.0	1187	5	PCT-US94-04496-28	Sequence 26, Appl
45	370.5	5.0	1210	1	US-08-320-559-26	

ALIGNMENTS

RESULT 1
US-09-103-429A-4
; Sequence 4, Application US/09103429A
; Patent No. 6187558
; GENERAL INFORMATION:
; APPLICANT: Granados, Robert R
; APPLICANT: Wang, Ping
; TITLE OF INVENTION: A No. 6187558el Invertebrate Intestinal Mucin
; TITLE OF INVENTION: CDNA and Related Products and Methods
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Pinnisi & Michaels, P.C.
; STREET: 118 No. 6187558th Tioga
; CITY: Ithaca
; STATE: NY
; COUNTRY: USA
; ZIP: 14850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/103.429A
; FILING DATE: 24-JUN-1998
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Michaels, Christopher A
; REGISTRATION NUMBER: 34,390
; REFERENCE/DOCKET NUMBER: BTI-39
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (607) 256-2000
; TELEFAX: (607) 256-3628
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 805 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Trichoplusia ni
; TISSUE TYPE: peritrophic membrane
; US-09-103-429A-4

Query Match 7.0%; Score 522; DB 4; Length 805;
Best Local Similarity 27.8%; Pred. No. 7.2e-25;

Mon Apr 29 08:35:43 2002

Db 958 KEEKAKK-----PEEKPKTEAKAKEDDKTISKPSKPKAEKAKSSSTDQKDSKPPKAT 1012

Qy 866 EN 867
|:
Db 1013 ED 1014

Search completed: April 26, 2002, 16:19:21
Job time: 346 sec

```
DR PIR; S00979; QFHUH.
DR MIM; 162230; -.
DR InterPro; IPR001664; IF.
DR Pfam; PF00038; filament; 1.
DR PROSITE; PS00226; IF; 1.
KW Intermediate filament; Coiled coil; Neurone; Phosphorylation.
FT DOMAIN 1 100
FT DOMAIN 101 413
FT DOMAIN 414 1020
FT DOMAIN 101 132
FT DOMAIN 133 145
FT DOMAIN 146 244
FT DOMAIN 245 266
FT DOMAIN 267 288
FT DOMAIN 289 292
FT DOMAIN 293 413
FT SEQUENCE 1020 AA; 111780 MW; 1177C9DCB3DCf1dA CRC64;

Query Match 6.8%; Score 502; DB 1; Length 1020;
Best Local Similarity 26.7%; Pred. No. 3.4e-13;
Matches 209; Conservative 81; Mismatches 356; Indels 136; Gaps 35;

Qy 119 KTKKVVIESE--EITEHSVS-ENQESSSSSSSSSSSTIWKIKSSKNSAANRELQKLLK 175
Db 336 KSTKDSLERQSELEDRHQADIASYQEAIOQLDAELRNKWM-----AAQLREYQDLIN 390
Qy 176 VK-----DANKNRKKKPTKPPVVDEAGSLDNGDFKVTTPTDSTTOHKK 221
Db 391 VKMALDIEIAAYKLLGECECRIGFGPIP-----FSLPEGLPKI--PSVST--HIK 437
Qy 222 VSTSPKITTAKPINRPSLPNSDTSKSLTVNKVETVETKETTNNKQISTDGKE-KT 280
Db 438 VKSEEEKIKVVE-----KSEKETVIVEE-----QTEETQVTEEVTEEEKEAKE 480
Qy 281 TSAKETQSIEKTSAKDLAPTISKVLAKPTPKAETTTKGPALTTPKRPPTTPKEPASTTPK 340
Db 481 EECGEEGEGEEEAEG-----GEEETKSP-----PAEEAASPEKEAKSPYKE 522
Qy 341 EPTPTTIKSAPTTPKPAPTTTKSAPTTTPKEPAPTTTKPAPTTTKPAPTTTKPAPTT 400
Db 523 EAKSPAEEKSPKEEAKSPAEEKS-PEKAKSPA-----KEEAKSPPEAKSPEKEEAKSPA 577
Qy 401 TKSAPTTPKPAPTTPKKAPTTPKPAPTTPKPAPTTPKPAPTTPKPAPTTPKPAP 460
Db 578 VKS-PEKAKSPAEEKAKSPA-----EAKSPEKAKSPVKEEAKSPA-EAKSPVKEEAKSPA 631
Qy 461 TAPKAPTTPKPAPTTPKPAPTTTKPSPPTTKPAPTTTKSAPTTTKPAPTTTKS 520
Db 632 VKSPEKAKSPKAEKSPKAEKSPKAEKSPKAEKSPKAEKSPKAEKSPKAEKSPKAEK 690
Qy 521 APTTPKEPSPTTTKPAPTTPKPAPTTPKKAPTTPKPAPTTPKPAPTTTKKPAPTA 580
Db 691 -PEKAKSP----VKEEAKSPKAEKSPVKEEAKSPV-----KEEAKTPKAKSPV- 740
Qy 581 PKPAPTTPKPAPTTTKKLTPTTPEKLTPTTPE--KPAPTTPEEL-APTTPPEPTTTP 637
Db 741 -KEEAKSPEKAKSPEKAKSLDVKSPEAKTFAKEARSFADKFEKAKSPVKEEVKSPKA 799
Qy 638 EEPAPTTPKAAAPNTPKPAPTTTP-----KEPAPTTPKPAPTTPKPAPTTTKGTAPTIL 693
Db 800 KSLPKADAKAPEKEIPKKEEVKSPVKEEKPQEVKVPKPAEKEKAPATPK----- 852
Qy 694 KEPAPTTPKKAPKELAPTTTKETSTTSKAPTTTKGTAPTTPKPAPTTTKPAPTT 753
Db 853 TEEKKDSKKEAPKKEAP-----KPKVEEKKEPAVEKESKVEAKKEADKKVP--T 905
Qy 754 PKGTAP--TTLKEPAPTTPKKAPKELAPTTTKGTSTTSKPA-PTTPKETAPTTPKEP 810
Db 906 PEKAPAKVEVKEDA-----KPKETEVAKKEPDADAKAKPEKSPKAEKKEAAP-EKKDT 957
Qy 811 APTTPKPKPAPTTTPPTTSEV--STPTTTKEPTTIHKSPE--STPELSAEPKPKAL 865
```


[2]
 RN SEQUENCE FROM N.A.
 RP MEDLINE-95347616; PubMed-7622068;
 RA Mita K., Tsuji H., Morimyo M., Takahashi E., Neno M.,
 RA Ichimura S., Yamauchi M., Hongo E., Hayashi A.;
 RT "The human gene encoding the largest subunit of RNA polymerase II.";
 RL Gene 159:285-286(1993).
 CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
 CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
 CC SUBSTRATES.
 CC -1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE -> N PROPHOSPHATE +
 CC RNA(N).
 CC -1- SUBUNIT: RNA POLYMERASE II CONSISTS OF 12 DIFFERENT SUBUNITS.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- PTM: THE TANDEN 7 RESIDUES REPEATS CAN BE HIGHLY PHOSPHORYLATED.
 CC -1- THE PHOSPHORYLATION ACTIVATES POL2.
 CC -1- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE
 CC FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA
 CC PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE
 CC III FOR 5S AND TRNA GENES.
 CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X63564; CA44125.1; -
 CC EMBL; X74874; CA52862.1; -
 CC EMBL; X74873; CA52862.1; JOINED.
 CC EMBL; X74872; CA52862.1; JOINED.
 CC EMBL; X74871; CA52862.1; JOINED.
 CC EMBL; X74870; CA52862.1; JOINED.
 CC EMBL; X74869; CA52862.1; JOINED.
 CC MIR; S21054; S21054.
 CC PIR; 180660; -
 CC InterPro: IPR000684; RNA_polII_repeat.
 CC InterPro: IPR000722; RNA_polA.
 CC InterPro: IPR002879; RNA_polA2.
 CC Pfam: PF00623; RNA_polA; 1.
 CC Pfam: PF01854; RNA_polA2; 1.
 CC PROSITE: PS00115; RNA_POLII_REPEAT; 43.
 CC Transferrase; DNA-directed RNA polymerase; Transcription; zinc; Repeat;
 CC DNA-binding; Nuclear protein; Phosphorylation; zinc-finger.
 CC 2N_FING 71 87 C2H2-TYPE (POTENTIAL).
 CC DOMAIN 1590 1958 CARBOXYL-TERMINAL 7-RESIDUE REPEATS.
 CC CONFLICT 1067 1067 W -> L (IN REF. 2).
 CC CONFLICT 1449 1449 D -> Y (IN REF. 2).
 CC SEQUENCE 1970 AA; 217205 MW; 6876FC25692A657E CRC64;

Query Match 6.8%; Score 503.5; DB 1; Length 1970;
 Best Local Similarity 33.3%; Pred. No. 5.1e-13;
 Matches 181; Conservative 101; Mismatches 167; Indels 95; Gaps 47;
 QY 326 PTPTTKEPASTTPE-PTPTTTSKAPTTPKEPAPTTPKAPTTPKE-PTPTTKE 379
 Db 1507 PSMGGISPAWTPWNGATPAYGAWSPVSGWTPGAGFSAASDASGFGSPGSPAWS 1566
 QY 380 PAPTTPKEPAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTP 437
 Db 1567 PTPGSGRGPSPYIPSPGGAMSPSYSPSTSPA-YEPRSPGTYTQSPSYSPSTSPSY-SP 1624
 QY 438 TTPKEPAPTTPKEPAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTP 497
 Db 1625 TSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY 1671
 QY 498 PAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTP 557
 Db 1672 -SPTSPSYSPSTSPSY-SPTSPSYSPSTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY 1725

QY 558 KEPAPTTPKEPAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTP 617
 Db 1726 SY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY 1777
 QY 618 PTTPEELAPPTPEPTPTTPEEPAPTTPKAAANTPKAPAPTTPKAPAPTTPKAPAPTTPKAPAPTTP 675
 Db 1778 PTSP-NYSPSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY 1828
 QY 676 TPKEAPTTPKGTAPTTPKAPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTP 734
 Db 1829 SP-SYSPSPSKYT-----PTSPSYSPSSPEYPTSPK--YSPTSPKYSPTSPK-YS 1875
 QY 735 PTPPKAPAPTTPKAPAPTTPKGTAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTP 793
 Db 1876 PTPSTY-SPTTPKY-SPTSP--TYSPTSPVYVTPSPK-----YSPTS---PTYSPTSPK 1922
 QY 794 PAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTP 851
 Db 1923 YSTSP-TYSPTSPKSTYSPTSPGY-SPTSP-----TYSLTSPAI-----SPDD 1965
 QY 852 STPE 855
 Db 1966 SDEE 1969
 RESULT 15
 NFH_HUMAN STANDARD; PRT; 1020 AA.
 AC P12036;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE NEUROFILAMENT TRIPLET H PROTEIN (200 KDA NEUROFILAMENT PROTEIN)
 DE (NEUROFILAMENT HEAVY POLYPEPTIDE) (NF-H).
 GN NEFH OR NEFH.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88328981; PubMed=3138108;
 RA Lees J.F., Sheldman P.S., Skuntz S.F., Carden M.J., Lazzarini R.A.;
 RT "The structure and organization of the human heavy neurofilament
 RT subunit (NF-H) and the gene encoding it.";
 RL EMBO J. 7:1947-1955(1988).
 CC -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
 CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
 CC NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT
 CC SUBSERVED BY THE TWO SMALLER NF PROTEINS.
 CC -1- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NPH IS
 CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
 CC THOUGHT THAT PHOSPHORYLATION OF NPH RESULTS IN THE FORMATION OF
 CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
 CC OF AXONAL CALIBER.
 CC -1- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
 CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H). THE
 CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
 CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
 CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X15306; CAA33366.1; -
 CC EMBL; X15307; CAA33366.1; JOINED.
 CC EMBL; X15308; CAA33366.1; JOINED.
 CC EMBL; X15309; CAA33366.1; JOINED.

327	TPTTPKPASTTKEPRTPTIKASAPTTPKAPATTTKSAPTTKEBPATTTTKEBPATTPPK 386
52	-----AAATAAEFTAAAGAPATTTTAPATTAAGKAPTAAATAPTAAAGAPTAT 103
387	EPATTTTKEBPATTTKS-APTTTKEBPATTPKKPAPATTTKEBPATTPKEPRTTTPKEPA- 444
104	GKAFATAAAMPVPTTAASKAPTAAATHSTAAAAAFTTAASAASKERSTSSSEEHCH 163
445	--PTTKBPAT--TTPKEBAPTAKKPATTPK-----EPATTPKBPATTPKE 489
164	VKPSKREMGSGKITKKQ----CKKNCCCPDKPGHGIIHCFHRKPKGHSHGHEHTTTTTK- 218
490	PSPTTPKEBPATTTKSAPTTTKEBPATTTKSAPTTTKEPSPTTTPKEBPATTPKEBPATTP 549
219	-----APTTIOIATTT--TPPTT-----TTTKATPTT----- 245
550	KKPAPTTPEBPATTPKEPATTTTKBPARTAPKEBPATTPKETAPTTPKKLPTTPEKLA 609
246	-----ITTUKATPTTT-----TTTKATTTT--TPTT----- 270
610	PPTPEKBPATTPBEELAPTTPEEPTPTTPEBPATTPKAAANPTKEBPATTPKEBPATTP 669
271	-----TTTTTKATPTTTTTTTT-----TTTTTKATPTTTTTTTT----- 289
670	KEBPATTPKEBAPTTPKGATPTTLKEBPATTPKKPAPKELAPTTTKEPTSTSDSKPAPT 729
290	-----TTTTTKATTTTTTSGECKMEPSK-----RDCGYSGITESQCR 328
730	PKG-----TAPPT-----PKBPATTPKEBPATTPKGATPTTLKEBPATTPKKPAPKELA 779
329	TKGCCFDSSIPOTKWCFYTLSQADCKVEQSQRVDCFRGIT----ADQCQKNCCEFS 384
780	PPTTKGPSTTSDSKPAPTTPKETAPTTKEBPATTPKKKAPATTPPETPPPTTSEVSPTTT 839
385	ISGTWKCFYSTSQVA--TKITPTPTTTTTTTTTTKATTTT-----TTTTTPTTT 436
840	KEPTTIHKSPDESTPELSAEPPTKALENSPKEGVPTTKTPAATKPEMTITAKDKTERD 899
437	TTTTT-----TTKA-----TTTTPTTTTTTTTTT---KAT--- 463
900	LRTPTTTTAAAPKMKEATTTTEKTESKITATTTQVSTTTQDTTPTFKITTLKTTTLAP 959
464	-TTTTPTTTT-----TTTTTKAT-----TTTTPTTTTTPTTTTKATTTPTTTTT 510
960	KVTITTKKTTTTEIMNKPETAAPKDRATNSKATTPKQKTPKAPKPTSTKKPK---T 1015
511	TTTTTKATTTTTSGECKME----PSKRADCGYPGITESQCRSKGCCEFSSIPOTKWCFYS 566
1016	MPRVKPKTTPTR 1029
567	LPOVDACKVAPSSR 580
RESULT 14	
RBPBI_HUMAN	STANDARD; PRT; 1970 AA.
ID	RBPBI_HUMAN
AC	P24928;
CD	01-MAR-1992 (Rel. 21, Created)
DT	01-MAR-1992 (Rel. 21, Last sequence update)
DT	20-AUG-2001 (Rel. 40, Last annotation update)
DE	DNA-DIRECTED RNA POLYMERASE II LARGEST SUBUNIT (EC 2.7.7.6) (RPB1).
DN	POLR2A.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RA	SEQUENCE FROM N.A.
RX	MEDLINE=92178992; PubMed=1542581;
RY	Wintzgrith M., Acker J., Vicaire S., Vigneron M., Keding C.;
RT	"Complete sequence of the human RNA polymerase II largest subunit.";
RL	Nucleic Acids Res. 20:910-910(1992).

InterPro: IPR002965; P-rich_extensn.
 DR PRINTS: PR01216; ADHESIVEI.
 DR PRINTS: PR01217; PRICHEXTENS.
 KW SIGNAL: Repeat; Hydroxylation.
 FT CHAIN: 1 20 POTENTIAL.
 FT DOMAIN: 21 872 ADHESIVE PLAQUE MATRIX PROTEIN.
 FT DOMAIN: 21 41 NONREPEITIVE LINKER.
 FT DOMAIN: 124 872 TANDEM REPEATS OF Y-K-[PS]-K-[IP]-[ST]-Y-P-[PST]-[ST].
 FT DOMAIN: 184 192 NONAPEPTIDE 1.
 FT DOMAIN: 213 221 NONAPEPTIDE 2.
 SQ SEQUENCE 872 AA; 101677 MW; 98CC70D7C75FF3C4 CRC64;

Query Match 7.0%; Score 518.5; DB 1; Length 872;
 Best Local Similarity 29.2%; Pred. No. 6.9e-14;
 Matches 276; Conservative 80; Mismatches 407; Indels 181; Gaps 54;

QY 194 VVDEAGSLDNGDKVTPDST--TQH-----NKVSTSKLITAKPINRPSLPNS 244
 DB 29 VYGSAYSGASAGAYK-TLPGSHYGSKHPVYKPMNKIPT-PYI--SKSYAPYKPKGY 84
 QY 245 DTSKETSLVNVKETTVEKTTTINKOTSTGKEKTTSAKETQSTIEKTSAKDLAPTSKVL 304
 DB 85 YPKRYOPTYGSNTYPPYKPIAKLSSYKAITYPAYKAKTSYPPSYK-----HKIT 139
 QY 305 AKPTKRAETTKGPAITPKPT-PTTPKEPASTTPKEPTPTTIKSAPTTPKEPATTTK 363
 DB 140 YPTTK-----PKITYP--PTYKQPSYPPSYKPTTYPPTYK-----PKITYPPTYK 185
 QY 364 SAPT-TPKEPAPT--TTKEPAPTTPKEPATTTPKEPATTTKSAPTTPKEPAPT--PKK 418
 DB 186 RKPSYTPYKATYPPYKPIIYP-----PTYKPKSYT-----PKKPTTYPPTYKPKI 236
 QY 419 PAPTTPKEPAPT--PKPEPTPT--PKPEAPTTPKEPAPTTPKEPAPT--APKAPAPTTP 471
 DB 237 SYSIYKPKASYSSYKSKKTYPTTKPKISYPTTKPKPSYP-----PTYKPKTYPPTY 292
 QY 472 KEPAPTTPKEPAPTTPKEPAPTTPKEPATTTKSAT-----TTKEPATTTKSAPTTPK 526
 DB 293 K-PKPSYPPYKPKITYPPY--YKPKSYPTPYKOKPSYPPYKSKSYPTSYKSKKTYPP 350
 QY 527 EPSPTTKEA-PTTPKEPAPTTPKP-APT-TPK-----EPAPTTPKEPAPTTPKPK 576
 DB 351 TYKPKIYPTTKPKPSYPPSYKPKTYPTTKPKITYPTTKPKPSYPPSYKPKTYPP 410
 QY 577 APTAPKEPAPTTPKETAP-----TTPKKLPTTPKEAPTTTPKEPAPTTPKEAPTTPE 631
 DB 411 T-YKPKISYPTTKPKASYSSYKSKKTYPTTKPKISYPTTKPKPSYPTTKPKITYP 469
 QY 632 PT-PTTPEEAPTTPKAAAPTTPK-EPAPTTPKEPAPTTP-----KPAPTTP--PKETAP 682
 DB 470 PTYKPKSYPTTKPKITYPTTKPKSYPTTKPKSYPTTKPKSYPTTKPKSYPTTKPK 529
 QY 683 TT--PKGTAPTTPKEPAPTTPKPKAPKEAPTTTKEPTSTTSKPAPT--TPKGTAPTTPK 739
 DB 530 PTYKPKITYPTTK-----PKPSYPPSYKPKITYPTTKPKIRYPPTTKPKASYPTTK 583
 QY 740 EPAPTTPKEPAPTTPKGTAPTTPKEPAPTTP-----KKAPKEAPTTTTPKGTSTTSKPA 795
 DB 584 -----PKITYPTTK--PKSYPTTKPKSYPTTKPKSYPTTKPKSYPTTKPKSYPTTK 628
 QY 796 PTT--PKETAPTTPKEPAPTTPK--KPAPTTPET-----PPPTSEVSTPTTKPEPT 844
 DB 629 PTKPKITYPTTK--PKPSYPPSYKPKITYPTTKPKPSYPPSYKPKSYPPSYKPK 687
 QY 845 IHKSPDESTPELSAPTTPKALENSKPEGVTTTPKATKEMTTAKDKT-----TERDL 900
 DB 688 Y---PPSYKPKKTYPTTK-----PKISYPTTK-----TKPSYAPSKKTSYPTTKPK 735
 QY 901 RTTPETTTAAKPMKETAATTTTEKTTESKITATTTQVSTTTTQDTTPPKITLTKTLLAPK 960
 DB 736 ISYPTTKAKP-----SYPTTKPKPSYSSYKPKIRYPPTTKPKPSYSSYK-----PK 785

QY 961 VTTTKTITTTTTEIMNKPEETAKPKDRATNSKATTPKQPQ---KFTKAPKP-TSTKPK-T 1015
 DB 786 I-----RYPPTYKPKP---SYASSYKPKIRYPPTYKPKPSYASSYKPKIT 827

QY 1016 MPRVYKPKPTT-PTPRKMTSTMPELNPTSTIAEAMLOTTTRPNO 1057
 DB 828 YPPTYKPKISYPTTKPKITYPTTKPKISYPPAYKPKISYPSQ 871

RESULT 12
 NFH_RAT ID NFH_RAT STANDARD; PRT; 831 AA.
 AC P16884; Q63368;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE NEUROFILAMENT TRIPLET H PROTEIN (200 KDA, NEUROFILAMENT PROTEIN)
 DE (NEUROFILAMENT HEAVY POLYPEPTIDE) (NF-H) (FRAGMENT).
 GN NFH OR NFH.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 ON NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=89065087; PubMed=3143606;
 RA Breen K.C., Robinson P.A., Wion D., Anderton B.H.;
 RT "Partial sequence of the rat heavy neurofilament polypeptide (NF-H).
 RT Identification of putative phosphorylation sites.";
 RL FEBS Lett. 241:213-218(1988).
 RN [2]
 RP SEQUENCE OF 37-831 FROM N.A.
 RX MEDLINE=88309090; PubMed=2457365;
 RA Dautigny A., Pham-Dinh D., Roussel C., Felix J.M., Nussbaum J.L.,
 RA Jolles P.;
 RT "The large neurofilament subunit (NF-H) of the rat: cDNA cloning and
 RT in situ detection.";
 RL Biochem. Biophys. Res. Commun. 154:1099-1106(1988).
 RN [3]
 RP SEQUENCE OF 1-89 AND 243-313 FROM N.A.
 RX MEDLINE=87080760; PubMed=2878828;
 RA Robinson P.A., Wion D., Anderton B.H.;
 RT "Isolation of a cDNA for the rat heavy neurofilament polypeptide
 RT (NF-H).";
 RL FEBS Lett. 209:203-205(1986).
 RN [4]
 RP SEQUENCE OF 318-831 FROM N.A.
 RX MEDLINE=89184647; PubMed=2928342;
 RA Lieberburg I., Spinner N., Snyder S., Anderson J., Goldgaber D.,
 RA Smulowitz M., Carroll Z., Emanuel B.S., Breitner J., Rubin L.;
 RT "Cloning of a cDNA encoding the rat high molecular weight
 RT neurofilament peptide (NF-H): developmental and tissue expression in
 RT the rat, and mapping of its human homologue to chromosomes 1 and
 RT 22.";
 RL FEBS Lett. 209:203-205(1986).
 CC -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
 CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
 CC NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT
 CC SUBSERVED BY THE TWO SMALLER NF PROTEINS.
 CC -1- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPLET K-S-P, NFH IS
 CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
 CC THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
 CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
 CC OF AXONAL CALIBER.
 CC -1- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
 CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
 CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
 CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
 CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
 CC -1- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 783
 CC ONWARD AND IS LONGER DUE TO A FRAMESHIFT.

Proc. Natl. Acad. Sci. U.S.A. 90:1536-1540(1993).

- FUNCTION: MIGHT FUNCTION AS A CALCIUM-SEQUESTERING "SPONGE" TO REGULATE THE AMOUNT OF FREE CYTOPLASMIC CALCIUM. IT BINDS 0.3 MOL OF CA+2 PER MOL OF PROTEIN.

- SUBUNIT: HOMODIMER (PROBABLE).

- SUBCELLULAR LOCATION: CYTOPLASMIC; HYPODENSE COMPARTMENT.

- TISSUE SPECIFICITY: SOMA AND AXONS OF PHOTORECEPTOR CELLS OF COMPOUND EYES AND OCELLI.

- DEVELOPMENTAL STAGE: EXPRESSED EARLY IN PHOTORECEPTOR CELL DEVELOPMENT.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; L02111; AAA28405.1; -;
EMBL; L05080; AAA28420.1; -;
PIR; A47282; A47282.
FlyBase; FBgn0010218; Cpn.
Calcium-binding.
KW CONFLICT 36 36 A -> AVAPAVVA (IN REF. 2).
FT CONFLICT 43 43 I -> T (IN REF. 2).
FT CONFLICT 64 64 I -> V (IN REF. 2).
FT CONFLICT 76 76 P -> A (IN REF. 2).
FT CONFLICT 100 100 T -> PP (IN REF. 2).
FT CONFLICT 126 127 VQ -> AP (IN REF. 2).
FT CONFLICT 154 154 I -> V (IN REF. 2).
FT CONFLICT 160 160 S -> T (IN REF. 2).
FT CONFLICT 534 534 A -> E (IN REF. 2).
FT CONFLICT 699 699 I -> T (IN REF. 2).
FT CONFLICT 703 703 V -> L (IN REF. 2).
FT CONFLICT 721 721 D -> E (IN REF. 2).
SQ SEQUENCE 865 AA; 84781 MW; 21104117E0B0E7CFE CRC64;

Query Match 7.2%; Score 530; DB 1; Length 865;
Best Local Similarity 26.2%; Pred. No. 2.5e-14;
Matches 233; Conservative 73; Mismatches 433; Indels 152; Gaps 41;

QY 298 APTSKVLAKP-TRK---AETTTKGPALTTPKEPTTPKEPASTTTPKEPTTPPTTIKSAPT 353
DB 8 SPVSAPVAAPVTSVAAPVQVVSRAVAP-----APAAPVAVTPVAPPPVTLASVQPAT 61
QY 354 PKEPAPTTKSAPTTP-KEPAPTTTKEPAPTTTPKEPAPTTTKEPAPTT-----TTKSAPTTP 408
DB 62 VTIPAPAPIAAASVTPVASVAPVVAAPT-----FAASPVSTPVAVAPVAVAPVAPV 116
QY 409 KEPAPTTPKKPAP-----TTKEPAPTTPKKEPT-----TTKEPAPTTKEPAPTT 454
DB 117 PVAATTPVQVIPAAPVATTPVVAASAPTAAVTPVISPVIASPPVVPVANTTPVAPV 176
QY 455 PKPEA-PTAPKKPAPTTKEPAPTTKEPAPTTTKE-PSPTTPKEPAP-----TTKSA 506
DB 177 AAVPAAPVAVPLAVAPVAVAPVVAEPVAPPPVAPVATPECVAPLPEVSVVATK 236
QY 507 PTTTKEP-----APTTTSAPTTKEPSPTTTPKEPA-PTTKEPAPTTPKKPAPTTTPKEPAP 562
DB 237 PLAAAEPPVVAAPATTPVVAASAPVSVAPAVETAVVAPVSVASTEPVVAALTTAP 296
QY 563 TTP-----KEPAPTTTTPKAPAPKEP-----APTTPKET-----APTTPKKLTPT 603
DB 297 ETPLAPVVAESQVAVANTVPTTPPAPEPETIAPPVVAETEVASVAVAEETTPVVPV 356
QY 604 TPCKL-APTTPKEPAPTTPEELAPTTPEPTTTPTEEP-----APTTPKAAAPTTPKEPAT 659
DB 357 AAESIPAPVVAITTPVAT---LAVTDPDVVASVAPVLPVIAAPVSPVSAVATPPVOLA 413
QY 660 TPKEPAPTTTPKEPAPTTTPKETAPTTP-KGTAPTTLKE-----PAPTTPKKPAPKE 708

Db 414 V-----LPPVAAEPVPAVVAEETPETPAPASAPVTIAALDIPEVAPVIAAPSDAPAEPSA 469
QY 709 LAPTTTREPSTSDSKPAPTPKGTAPTTP-----KEPAPTTPKEP---APTTPKGTAPT 760
Db 470 AAPIVSTPP--TTASVPETTPAPPAVTEPIDVSVLSEAAIETPVAPPVEVTTTEVAVADV 527
QY 761 TLKEPA-----PTTPKKPAPKELAPTTT-----KGTPTSTTSDKPAPTT-----PKETAPT 806
Db 528 APPEAAADLIIEPVPEPPAPIDLLLEQTTSVPVAAEASTSS--PIPETSLPPNEAVASP 585
QY 807 PKEPAPTTPKKPAPTTPEPTTPPTTSEVSTPTTKEPTTHKSPD-----ESTPEL 856
Db 586 EVAVAPITAPEPIP-----EPEPSLATPTTEIPVEAPVVIQEAADVAVPVTETSTIPET 641
QY 857 SAEPTPKALNSKPEKGVPTTKTAPATKPMTTTAKDKTKTTERDLRTP--ETTTPAAPT 914
Db 642 TVE-FPEAVAEKVLDPAL--TEAPVTTQEPDVANINDGAPATEI--TPPAVEIVTAAAEVS 697
QY 915 KETATTTKTESKITATTQVST-----TTQDTPPKITTLKTTTLA-PKVTTTKT 967
Db 698 DIAIPVIDPPVPOEIAVAEIPETDTKPAEIVREQSTIPEAPVPEVSKVAPVISEAPAA 757
QY 968 ITTTEIMNKPEETAKPKDRATNSKATTPKPKPTKAPKPTSTKPKTPRVRKPKPTTPT 1027
Db 758 EVPITAGDNDNTSGISEVPTIAEKPVVEPTS--EIQSSSSPS--DSVPVAKITPL 813
QY 1028 PRKMTSTMPELNPTSRIAM---LQTTTRPNQTPNSKLVEVN-----PKS 1070
Db 814 LRDLQTTDVSLAIAATLDAIGEKLKDKARKQOVMDRLCEIEKILGPPKS 864

RESULT 11
FPI_MYTGO
ID FPI_MYTGO STANDARD; PRT; 872 AA.
AC Q25434;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ADHESIVE PLAQUE MATRIX PROTEIN PRECURSOR (FOOT PROTEIN 1) (MCFP1).
GN FPI.
OS Mytilus coruscus (Sea mussel).
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloida;
OC Mytiloidae; Mytilidae; Mytilus.
OX NCBI_TaxID=42192;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=FOOT;
RX MEDLINE=96394686; PubMed=8798340;
RA Inoue K., Takeuchi Y., Takeyama S., Yamaha E., Yamazaki F., Odo S., Harayama S.;
RT "Adhesive protein cDNA sequence of the mussel Mytilus coruscus and its evolutionary implications";
RL J. Mol. Evol. 43:348-356(1996).
CC - FUNCTION: PROVIDES ADHESIVENESS TO THE MUSSEL'S FOOT. MUSSELS PRODUCE ONE OF THE STRONGEST WATER INSOLUBLE GLUE. THE MUSSEL'S ADHESIVE IS A BUNDLE OF THREADS, CALLED A BYSSUS, FORMED BY A FIBROUS COLLAGENOUS CORE COATED WITH ADHESIVE PROTEINS.
CC - SUBCELLULAR LOCATION: SECRETED.
CC - TISSUE SPECIFICITY: PRODUCED BY THE BYSSAL GLAND.
CC - DOMAIN: ALMOST EXCLUSIVELY COMPOSED OF REPEATS OF A DECAPEPTIDE.
CC - PTM: PROLINES IN THE REPEAT ARE HYDROXYLATED (SINGLE OR DI-) AND ALSO TYROSINE (THUS PRODUCING DOPA = 3,4-DIHYDROXYPHENYLALANINE).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; D63777; BAA09850.1; -;
InterPro; IPR002964; Adhesive_plaq.
DR

QY	327	TPPTTKE-PASTTKEPTTTIKSAPTTKEPAPTTTKSAPTTKEPAPTTTKKEPAPTTTKEPAPTTT	385
Db	599	TPSTPADSSAHSSTPSTPVDSS---SAHSTPSTPADSSAHSSTPSTPVDSSAHSST---PSTP	651
QY	386	KE-PAPTTTKEPAPTTTKSAPTTTKE-PAPTTKPKP---PAPTTKEPA---PPTP	432
Db	652	ADSSAHSSTPSTPVDSSAHSSTPSTPVDSSAHSSTPSTPVDSSAHSSTPSTPVDSSAHSSTP	711
QY	433	KEPTP-TTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP	491
Db	712	VDSSAHSSTPSTPADSSAHSSTPSTPADSS---AHSSTPADSSAHSSTPSTPVDSSA	764
QY	492	PTTKEPAPTTTTSAPTTTKEPAPTTTTSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP	547
Db	765	HGTPSTPADSSAHSSTPST---PADSSAHSSTPSTPVDSSAHSST---PSTPVDSSAHSSTP	817
QY	548	TP-KKPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT	602
Db	818	TPVDSSAHSSTPSTPVDSSAHSSTPSTPVDSSAHSSTPSTPVDSSAHSSTPSTPVDSSAHS	876
QY	603	TTP-EKLAPTTTKEPAPTTTPEELAPTTPEE---PTPTTPEPA---PTTP-KAAPNT	652
Db	877	STPVDSSAHSSTPSTPADSSAHS-TPTPVDSSAHSSTPSTPADSSAHSSTPSTPVDSSAHS	935
QY	653	PREPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE	712
Db	936	PSTPADSS---AHSSTPSTPADSSAHSST---PSTPADSS---AHSSTPSTPVDSSAHS	983
QY	713	TTKEPTSTTSDKAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE	771
Db	984	---PSTPADSSAHSST---PSTPADSSAHSSTPSTPADSSAHSSTPSTPVDSSAHSSTP	1033
QY	772	KPAKELAPTTTKEPAPTTTSDKAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT	827
Db	1034	TP-----ADSSAHSSTPSTPADSSAHSSTPSTPVDSSAHSSTPSTPADSS---AHSSTP	1084
QY	828	PTTSEVSTPTTKEPTTTHKSPDESTP-ELSAEPTPKALENSKPEGVPTTKTPAATKE	886
Db	1085	DSSAH-STPSTPAD-SSAHSSTP-STPADSSAHSSTP-----STPAGSSAN	1125
QY	887	MT 888	
Db	1126	GT 1127	
RESULT	10		
CPN_DROME	STANDARD;	PRT;	865 AA.
ID	CPN_DROME		
AC	Q02910;		
DT	01-OCT-1993 (Rel. 27, Created)		
DT	01-OCT-1993 (Rel. 27, Last sequence update)		
DT	01-FEB-1994 (Rel. 28, Last annotation update)		
DE	CALPHOTIN		
GN	CPN OR CAP.		
OS	Drosophila melanogaster (Fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
OC	Ephydroidea; Drosophilidae; Drosophila.		
OX	NCBI_TaxID=7227;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-CANTON-S;		
FX	MEDLINE=93165729; PubMed=8094559;		
RA	Martin J.H., Benzer S., Rudnicka M., Miller C.A.;		
RT	"Calphotin: a Drosophila photoreceptor cell calcium-binding protein."		
RL	Proc. Natl. Acad. Sci. U.S.A. 90:1531-1535(1993).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-CANTON-S;		
FX	MEDLINE=93165730; PubMed=8434015;		
RA	Ballingier D.G., Xue N., Harshman K.D.;		
RT	"A Drosophila photoreceptor cell-specific protein, calphotin, binds		
RT	calcium and contains a leucine zipper.";		
RT			

Query Match	7.2%	Score 530.5;	DB 1;	Length 1162;
Best Local Similarity	29.4%	Pred. No. 3e-14;		
Matches 177;	Conservative 100;	Mismatches 212;	Indels 113;	Gaps 34;

582	QY	KEPAPTTPKETAPT-----TPKKLITPTPEKLA-----PTTPEKPAPTTPEELAPTTPEE	631
583	QY	!!!:!!!	

OY 913 MKETATTEKTESKATATTTQVSTTTODTTPKITTLLKTTTLAPVTTTKTTTTE 972
DB 748 -TYKAKPPYPPYKAK-----PSYPPYKPKPSYKSKSIYSPSKPKTKTTP- 798
OY 973 INKKEDEAKPKDRATNSKATTPKQKPTKAKK---PTSTKPKTPMVRKKPKTTP- 1028
DB 799 --YKFKLTPPYK-----PKSPSYKPKKITYPSTYKTKLPSYPPYKSKTSKTSPT 848
OY 1029 -----RKMTSTPELNPTSR 1043
DB 849 YNKKISYSPSYKAKTSYPAKPKPTR 874
RESULT 6
ZAN_HUMAN
ID ZAN_HUMAN STANDARD: PRT; 2700 AA.
AC O9Y493; 000218;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ZONADHESIN (FRAGMENT).
GN ZAN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE OF 1-2379 FROM N.A.
RX MEDLINE-99018118; PubMed-9799793;
RA Glocker G., Scherer S., Schattevoy R., Boright A., Weber J.,
RA Tsui L.C., Rosenthal A.;
RT "Large-scale sequencing of two regions in human chromosome 7q22:
RT analysis of 650 kb of genomic sequence around the EPO and CUTL1 loci
RT reveals 17 genes.";
RL Genome Res. 8:1060-1073(1998).
RN [2]
RP SEQUENCE OF 2338-2700 FROM N.A.
RX TISSUE-Testis;
RA MEDLINE-97271566; PubMed-9126492;
RA Geo 2., Harumi T., Garbers D.L.;
RT "Chromosome localization of the mouse zonadhesin gene and the human
RT zonadhesin gene (ZAN).";
RL Genomics 41:119-122(1997).
CC -1- FUNCTION: BINDS IN A SPECIES-SPECIFIC MANNER TO THE ZONA PELLUCIDA
CC OF THE EGG. MAY BE INVOLVED IN GAMETE RECOGNITION AND/OR
CC SIGNALING.
CC -1- SUBUNIT: PROBABLY FORMS COVALENT OLIGOMERS.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN, EXCLUSIVELY ON THE
CC APICAL REGION OF THE SPERM HEAD (BY SIMILARITY).
CC -1- TISSUE SPECIFICITY: IN TESTIS, PRIMARILY IN HAPLOID SPERMATIDS.
CC -1- DOMAIN: THE MAM DOMAINS PROBABLY MEDIATE SPERM ADHESION TO THE
CC ZONA PELLUCIDA.
CC -1- DOMAIN: DURING SPERM MIGRATION THROUGH THE REPRODUCTIVE TRACTS,
CC THE MUCIN-LIKE DOMAIN MIGHT INHIBIT INAPPROPRIATE TRAPPING OF
CC SPERMATOZOIA OR PROMOTING ADHESION TO THE OVUDUCTAL ISTHMS.
CC -1- DOMAIN: THE VWFD DOMAINS 2 AND 3 MAY MEDIATE COVALENT
CC OLIGOMERIZATION (BY SIMILARITY TO HUMAN INTESTINAL MUCIN MUC2).
CC -1- SIMILARITY: CONTAINS 3.5 MAM DOMAINS.
CC -1- SIMILARITY: CONTAINS 4.5 VWFD DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF053356; AAC78790.1; -
CC EMBL; U83191; AAC51208.1; -
CC MIM: 603372; -
CC InterPro: IPR000561; EGF-like.
CC InterPro: IPR000998; MAM.

DR InterPro: IPR002965; P-rich_extensn.
DR InterPro: IPR002919; TIL.
DR InterPro: IPR003328; TILA.
DR InterPro: IPR001007; VMFC.
DR InterPro: IPR001846; Vwd.
DR Pfam: PF006529; MAM; 3.
DR Pfam: PF01826; TIL; 4.
DR Pfam: PF02345; TILA; 4.
DR Pfam: PF00094; Vwd; 4.
DR PRINTS: PR01217; PRICHEXTENS.
DR SMART: SM00137; MAM; 2.
DR SMART: SM00014; VNC; 1.
DR SMART: SM00016; VNC_def; 3.
DR PROSITE: PS01186; EGF_2; 3.
DR PROSITE: PS00740; MAM_1; 1.
DR PROSITE: PS50060; MAM_2; 4.
KW Glycoprotein; Transmembrane; Cell adhesion; Repeat.
FT NON_TER 1 1
FT DOMAIN <1 109 MAM 1.
FT DOMAIN 112 136 MAM 2 (PARTIAL).
FT DOMAIN 161 326 MAM 3.
FT DOMAIN 322 446 MAM 4.
FT DOMAIN 483 951
FT DOMAIN 953 1065
FT DOMAIN 1066 1454 VWFD 2.
FT DOMAIN 1455 1861 VWFD 3.
FT DOMAIN 1862 2292 VWFD 4.
FT DOMAIN 2293 2684 VWFD 5.
FT DOMAIN ? 2 EGF-LIKE.
FT CARBOHYD 74 74 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 403 403 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1023 1023 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1099 1099 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1618 1618 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1737 1737 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1832 1832 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1878 1878 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2136 2136 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2505 2505 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 2374 2379 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON_TER 2700 2700 NNOKMA -> RAGPGP (IN REF. 1).
SQ SEQUENCE 2700 AA; 293013 MW; 80E60CC0B12277B1 CRC64;
Query Match 7.5%; Score 555.5; DB 1; Length 2700;
Best Local Similarity 32.6%; Pred. No. 6,4e-15;
Matches 199; Conservative 64; Mismatches 240; Indels 107; Gaps 32;
OY 233 PINPRSLPPNSDTSKESLTVNKKETVETKTTTNNKOTSTDGKETSNAKQSIK 292
DB 445 PKVLPDLPPVSVSS-----TSPSETTGTEPNPTSTK-----KPTVSIK 487
OY 293 SAKDLAPTSKVLAKPTPAETTTKGPALTTPKEPTTPPKPEASTTPKEPT-----PTTI 347
DB 488 SVTTEKPT-----VPEKPKPTIPEKPTISTEK-----PTIPEKPPNPKSKPTIIL 539
OY 348 KCAPTPKEPAATTTKSAAPTTPKEPAATTPKEPAATTPKEPAATTKPA-PTTTSAPT 406
DB 540 TEKPTIPE-KPTIPEKPTISTEKPTVTEP-PTTPEETTYMEAPVPTPEKPT 596
OY 407 TPKEPAATTPKAPATTPKEPAATTPKAPPTTPKEPAATTPKAPATTPKAPATPK 466
DB 597 --EKSIPTK-----PTISMEETIISTEKPTICPEKPTIPEK-----PTIPEKSTISPEK 647
OY 467 APTTPKEPAATTPKAPATTPKESPTTPKPEAPATTPKSAPTTTKEPAATTPK 526
DB 648 -PTTPE-KPTIPEKPTISTEKPTIPEK-PTISPEKLTPTKLTPT- -EKPTIPT 701
OY 527 EPSPTTKEPAATTPKPEAPATTPKKA-----PTTPKEPAATTPKPEAPATTPK 581
DB 702 EKPTISTE-PTTPEETTTISTEKPSIPEKPTIPEETTTTVEETTTISTEKLITD- -757

```

OY      888 TTANKDTERDRLTPETT-----TAAPKMKKEATITEKTETSKIAITATTOYSTT 940
       | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      754 --PAGVTSAPDTRPAAGSTAPPRAHGVTSAP---DIRPAPGSTAPPRAHGVTSAPDTRPA 807
OY      941 TODTP--FKTTTKTTTLAKVTYTKKTIITT--EIMNMDEETAKEKDRAATNSKATTPK 996
Db      808 PGSTAPPRAHGVTSAPDTRRAPAGSTAPPRAHGVTSAPDTRRAPAGSTAPPRAHGVTSAPDTRPA 867
OY      997 POKPTKADKKRPTSTKKPKTPRVKRKPKTTPPRKMTSTMELNPTSKLAEMLOTTTRPN 1056
       | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      868 P--GSTAPPRAHGVTSAPDTRP--APGSTAPPRAHGVTSAPDTRPA-----PG 909
OY      1057 OT-PMSKIYEYNPKSEDAAGEENPHMLLRPHVEFMEPTPRDMOYLLEPVNOGIITNPML 1115
       | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      910 STAPPRAHGVTSAPDTRRAPPGSTAPPRAH-----GVTSAPDNRRALGSTA----PV 955
OY      1116 SDETNICNGKPYDGLTTLRNGTLVAFRGHYFWLMSPPSPPS 1156
       | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      956 HNVTSASGSASGASTLVNCHTSARATTTPASKSTPFSPIS 996

RESULT          5
FPL_MYTDED      STANDARD:      PRU;      875 AA.
ID             FPL_MYTDED
AC      025460:
DT      30-MAY-2000 (Rel. 39, Created)
DT      30-MAY-2000 (Rel. 39, Last sequence update)
DT      20-AUG-2001 (Rel. 40, Last annotation update)
DE      ADHESIVE PLAQUE MATRIX PROTEIN (POLYPHENOLIC ADHESIVE PROTEIN) (FOOT
GN      PROTEIN 1) (MEP1) (FRAGMENT).
DE      FPI.
OS      Mytilus edulis (Blue mussel).
OC      Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloidea;
OX      Mytiloidae; Mytilidae; Mytilus.
NCBI_TaxID=6550;
RN      [1]
RX      SEQUENCE FROM N.A. MEDLINE=91025829; PubMed=1367451; Strausberg S.L., Strausberg R.L.;
RT      Filipula D.R., Lee S.M., Link R.P., Strausberg S.L., Strausberg R.L.;
RT      "Structural and functional repetition in a marine mussel adhesive
RT      protein.";
RT      Biotechnol. Prog. 6:171-177(1990).
RL      [2]
RP      PARTIAL SEQUENCE, AND POST--TRANSLATIONAL MODIFICATIONS.
RX      MEDLINE=83135732; PubMed=6298211;
RA      Waite J.H.;
RT      "Evidence for a repeating 3,4-dihydroxyphenylalanine- and
RT      hydroxyproline-containing decapeptide in the adhesive protein of the
RT      mussel, Mytilus edulis L.";
RL      J. Biol. Chem. 258:2911-2915(1983).
CC      -1- FUNCTION: PROVIDES ADHESIVENESS TO THE MUSSEL'S FOOT. MUSSELS
CC      PRODUCE ONE OF THE STRONGEST WATER INSOLUBLE GLUE. THE MUSSEL'S
CC      ADHESIVE IS A BUNDLE OF THREADS, CALLED A BYSSUS, FORMED BY A
CC      FIBROUS COLLAGENOUS CORE COATED WITH ADHESIVE PROTEINS.
CC      -1- SUBCELLULAR LOCATION: SECRETED.
CC      -1- TISSUE SPECIFICITY: PRODUCED BY THE BYSSAL GLAND.
CC      -1- DOMAIN: ALMOST EXCLUSIVELY COMPOSED OF REPEATS OF A DECAPEPTIDE.
CC      -1- PTM: THE DECAPEPTIDE A-K-P-S-Y-P-P-T-Y-R IS POST-TRANSLATIONALLY
CC      MODIFIED AS FOLLOWING: THE SIXTH AND SEVENTH RESIDUES ARE
CC      HYDROXYLATED AND THE PENULTIMATE IS A 3,4-DIHYDROXYPHENYLALANINE
CC      (DOPA) DERIVED FROM TYROSINE.
CC      -----
CC      THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation at
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL, X54422; CA38294.1; -.
DR      InterPro: IPR002964; Adhesive_plaq.
DR      InterPro: IPR002965; P_rich_extensin.

```

[illegible]

Query Match	Best Local Similarity	8.8%	Score 651;	DB 1;	Length 1355;
Matches 295;	Conservative 72;	Mismatches 460;	Indels 234;	Gaps 51	
FT CARBOHD 1133 1133	N-LINKED (GLCNAC...) (POTENTIAL).				
FT VARSPLC 19 19	T -> TATTAPEPAT (IN ISOFORM B).				
FT VARSPLC 20 22	MISSING (IN ISOFORM C).				
FT VARSPLC 20 31	MISSING (IN ISOFORM D).				
FT VARSPLC 126 905	MISSING (IN POLYMORPHIC EPITHELIAL ISOFORM).				
FT VARSPLC 1077 1087	FLQIKGGFL -> VSLGSLPMLP (IN SECRETED ISOFORM).				
FT VARSPLC 1088 1255	MISSING (IN SECRETED ISOFORM).				
FT CONFLICT 2	T -> A (IN REF. 11).				
FT CONFLICT 134 134	P -> Q (IN REF. 9).				
FT CONFLICT 154 154	S -> Q (IN REF. 9).				
FT CONFLICT 1021 1021	S -> T (IN REF. 3).				
FT CONFLICT 1251 1251	A -> T (IN REF. 3).				
FT SEQUENCE 1255 AA; 122072 MW; 5E28DFC4DEFD9A82 CRC64;					
QY 250 TSLVNNKETVEIKETTTNKTSTDGKEKTSAKETOSIEKTSAKD-LAPTISKVLAKPT 308					
Db 16 TLVLV-----VTGSGHASSIRPGGKEKLSAQRSSVSPSTEKNVSMSSVLSHNS 65					
QY 309 P-KATITTKGP--ALTTPKEP-----TPPTTKEPASTTPKEPPPTT--K 348					
Db 66 PGGSGSTTGQGDVTLAPATEPASGSAATWGODVTSVPYTRPALGSTTPPHANDVTSAPDNK 125					
QY 349 SAPPTTKEPAPTTTKASAPTTKEPAATTTTKEP-----APPTKKEAPTTTKEPAPTTTK 402					
Db 126 PAFGSTAPRAHGVTSAPDT--RAPGSTAPRAHGVTSAPDT--RAPGSTAPRAHGVTSAPDT-- 179					
QY 403 SAPPTTKEPAPTTT-----KRAPPTTKEP-----APPTTKKEPTPTT----- 440					
Db 180 SAPDTRRAPGSTAPRAHGVTSAPDTRRAPGSTAPRAHGVTSAPDTRRAPGSTAPRAHGVTS 239					
QY 441 -----KEAPPTTKEP-----APPTTKKEAPPTTK-----KRAPPTTKKEAPPTTK-----E 481					
Db 240 SAPDTRRAPGSTAPRAHGVTSAPDTRRAPGSTAPRAHGVTSAPDTRRAPGSTAPRAHGVTS 299					
QY 482 PAPPTTKEPSTPTTKEPAPTT--TKSAPTTTKKEAPTTTKSAPPTTKKEPSTPTTKEP-- 536					
Db 300 SAPDTRRAPGSTAPRAHGVTSAPDTRRAPGSTAPRAHGVTSAPDT--RAPGSTAPRAH 356					
QY 537 -----APPTTKKEAPPTTK-----KRAPPTTKKEAPPTTK-----EPAPTTTKKRAPPTTKEP 584					
Db 357 GVTSAAPDTRRAPGSTAPRAHGVTSAPDTRRAPGSTAPRAHGVTSAPDTRRAPGSTAPRAH 416					
QY 585 APPTTKKEAPTTTKKLTPTTPEKLAPTTPEKAPPTTPEKLAPTTPEEPTPTTPEEAPPTT 644					
Db 417 GVTSAAPDTRRAPGSTAPRAHGVTSAPDTRRAPGSTAP--PAHGVTSAPDTRRAPGSTA 472					
QY 645 PKA-----AANTPTKEAPPTTK-----EPAPTTTKEPAPTTTKETAPTTKAPTTTKEP 696					
Db 473 PRAHGVTSAPDTRRAPGSTAPRAHGVTSAPDTRRAPGSTAP--PAHGVTSAPDTRRAP 528					
QY 697 APPTT-----KKPAPEKLAPTTKEPTSTSDPKAP--TPPKGATPTTKEAPPT 744					
Db 529 GSTAPRAHGVTSAPDTRRAPGSTAP--PAHGVTSAPDTRRAPGSTAPRAHGVTS--APD 583					
QY 745 TKPEEAPPTTKG-----TAPTTTKKEAPDTT-----KKRAPKELAPTTTKGPTSTT 790					
Db 584 TRAPGSTAPRAHGVTSAPDTRRAPGSTAPRAHGVTSAPDTRRAPGSTAP--PAHGVTSAP 642					
QY 791 SDKPAP--TPPK-----ETAPPTTKKEAPPTT-----KKRAPPTTPEPTPTTSEVS 834					
Db 643 DTRAPGSTAPRAHGVTSAPDTRRAPGSTAPRAHGVTSAPDTRRAPGS--TAPRAHGVTS 700					
QY 835 TPTTTEKPTT-----IHKSPESTPELSAEPTRKALENSPKPEKGVPTTKTPATGPEM 887					
Db 701 APDTRRAPGSTAPRAHGVTSAPDTRRAPGSTAPRAHGVTSAP-----DTRAPGSTAP-- 753					


```

OY 810 PAPTPKPKAPV-TPETPPPTSEVSTPTTKKEPTTIHKSPDESPTELSAETPKALENS 868
DB 1254 PTPS--DEPTPSEDEEPIPTDTPSEDEPTPSEPT--PSDEPTP--SDEPTPSEDEPT 1305
OY 869 PREPVPPTTKPAAT--PREMTTAKDKTTEDELTATPETTAAPKMKKEATTTTEKTE 926
DB 1306 SDEPTPSEDEPTPSEPTPIPTDTPSEDEPTSD--EPTEPSEDEPTPSEDEPTP 1361
OY 927 SKITATTTTQVSTTTQDTTPPKITTLKTTLTAPKVTTKKITTTEIMNKPEETAKPKDR 986
DB 1362 DEPTPSEDEPTPSEPTPEEPTP--TTTPPTPSTT-----PTSG 1366
OY 987 ATNSKAT-----TPPQKPTKAP--KKPTSTKKPKTPMPVKPKKTTTPPKM----- 1031
DB 1397 SGGSGSGSGGGGGGGGCTVPSTPTPTPSKPTSTPAP--TEIEPTPSDVGAIGEHRA 1453
OY 1032 -----TSTMPLEPTPSRIAEMIQTTTRPNQPNPSKLVNKSDEAGAESEPTPM- 1083
DB 1454 YLRGYPDSFRPERNITRAEAAYF-----AKLL-----GADESYCAQASAPYSD 1498
OY 1084 LLRPH-----VEMPE-----VTPMDYLPRVNOGI----- 1109
DB 1499 LADTHAMAMAKFATSOGLFKGYDPGTFKPDQNTIRAEFATVVLHFLTKVKGQIMSKLA 1558
OY 1110 ---TIINPLSDETNICNG---KPVDELTL-----RNGTLVAFRGHYTMLSPEPSPS 1156
DB 1559 TIDISNKRFPD---CVGHMAQEFIEKLTSLGYISGYPDST-----FKPON 1600
OY 1157 PARRITEWGIPEPIDVTFRNCEGKTFPFKD--SQWTFPTDIXD 1201
DB 1601 YIKRESV---ALINRALEKRGPLNGAPKLPDVNESYMAF-GDIMD 1642

```

RESULT 3

```

ID AMYL_YEAST STANDARD: PRT: 1367 AA.
AC P08640; P08068;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1995 (Rel. 31, last sequence update)
DT 20-AUG-2001 (Rel. 40, last annotation update)
DE GLUCAMYLASE SI/S2 PRECURSOR (EC 3.2.1.3) (GLUCAN 1,4-ALPHA-
GN GLUCOSIDASE) (1,4-ALPHA-D-GLUCAN GLUCOHYDROLASE).
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Bartell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA Churcher C.M., Connor R., Corsey T., Dear S., Devlin K., Fiser A.,
RA Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
RA Louis E., Lye G., Moule S., Moule T., Odeil C., Pearson D.,
RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA Walsh S.V., Whitehead S.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-242 AND 762-1331 FROM N.A.
RX MEDLINE=87194600; PubMed=3106330.
RA Yamashita I., Nakamura M., Fukui S.;
RT "Gene fusion is a possible mechanism underlying the evolution of
RT STRA1."
RL J. Bacteriol. 169:2142-2149(1987).
RN [3]
RP SEQUENCE OF 1-31 FROM N.A.
RC STRAIN=SPX101-1C;
RX MEDLINE=89031230; PubMed=3141213;
RA Pardo J.M., Ianez E., Zalacain M., Clares M.G., Jimenez A.;
RT "Similar short elements in the 5' regions of the STRA2 and SGA genes
RT from Saccharomyces cerevisiae."
RL FEBS Lett. 239:179-184(1988).

```

```

CC -|- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL 1,4-LINKED ALPHA-D-
CC GLUCOSE RESIDUES SUCCESSIVELY FROM NON-REDUCING ENDS OF THE CHAINS
CC WITH RELEASE OF BETA-D-GLUCOSE.
CC -|- SIMILARITY: TO S.POMBE SPC215.13.
CC -|- SIMILARITY: SOME, TO S.POMBE SPC285.13C.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb-sdb.ch).
CC
DR EMBL: Z38061; CAA86176.1; -
DR EMBL: M16164; AAA35014.1; -
DR EMBL: M16165; AAA35015.1; -
DR EMBL: X13857; CAA32069.1; -
DR PIR: B26877; B26877.
DR PIR: A26877; A26877.
DR PIR: S48478; S48478.
DR SGD: S0001458; MDC1.
KW Hydrolyase; Glycosidase; Polysaccharide degradation; Glycoprotein;
KW Signal; Multigene family.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 1367 GLUCAMYLASE SI/S2.
FT DOMAIN 210 1367 SER/THR-RICH.
FT CARBOHYD 817 817 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 874 874 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1367 AA; 136110 MW; 91C00E2DBD61A9D CRC64;

```

Query Match 10.9%; Score 809; DB 1; Length 1367;
 Best Local Similarity 27.5%; Pred. NO. 6.1e-25;
 Matches 328; Conservative 112; Mismatches 526; Indels 226; Gaps 53;

```

OY 61 DCDAAQKKYDK---CCPYDESEFCAEYHNPTSPSSKKAPPPGASOLTKTKRKPKN 117
DB 174 DLSTGCNNYDNGHSGQTDPRGFYWNIDCNGCGTKSSITTSSES--STT----- 223
OY 118 KKKTKKVESEITTEHSYSENOESSSSSSSSSTTWIKSKNSANRELQKLVK 177
DB 224 ---TSTSESTSTTSTSESSSTTSTSESSSTT-----APATPTTSCTEKEPT 274
OY 178 DNKKRRTKKKPKPKPVVDEAGSGLDNDGFKVTPDSTGTOHNKVSSTPKTTAKPIPR 237
DB 275 PTTTSCTEKEPT--PPHND-----TTPCT-----KKKTTYSKTCRKTTPV 314
OY 238 PSLPNSDTSKETSILVNKEETVETKETTNNKOTSTDGKEKTSAKETOSIEKTSADKL 297
DB 315 PT--PSSSTSESS-----APVPTPSSST-----ESSSAPVTSSTSESS-- 353
OY 298 APTSKVLAKPTPKKETTNGKPAITPKPEPTTTPKREPASTTPKKEPTTTPKSAPTTREP 357
DB 354 APV-----PTPSSSTSESSAPVTSSTSESSAPVTSSTSESSAPVTPPSSSTSESS 407
OY 358 APTT-----TKSAP-----TPKREPAPTTKREPAPTTKREPAPTTKSA 404
DB 408 APVTSSTSESSAPVTSSTSESSAPVTS-----TTSESSAPVTSSTSESSAPVTPPSS 463
OY 405 PTPKREPAPTTPKKREPAPTTKREPAPTTKREPAPTTKREPAPTTKREPAPTTKREPAP 461
DB 464 STTESSAPVT---SSTSESSAPV---PTPSSSTSESSAPVT---SSTSESSAPV 512
OY 462 APKKPAPTTKREPAPTTKREPAPTTKREPAPTTKREPAPTTKREPAPTTKREPAPTTKSA 521
DB 513 -----PTPSSSTSESSAPVTPS---SSTSESSAPVTPS---TTSESSAPVTPPSS 559
OY 522 PTPKESPTTKKREPAPTTKREPAPTTKREPAPTTKREPAPTTKREPAPTTKREPAPTTK 580
DB 560 STTESSSTPVTSS---TTSESSAPVTPPSSSTSESSAPVTP---PSSSTSESSAPV 611
OY 581 PKREPAPTTKREPAPTTKREPAPTTKREPAPTTKREPAPTTKREPAPTTKREPAPTTKREP 639

```


TT	CARBOHYD	1787	1787
TT	CARBOHYD	1820	1820

[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:19:03 ; Search time 47.4 Seconds
(without alignments)
1067.457 Million cell updates/sec

Title: US-09-556-246-1_COPY_25_1404

Perfect score: 7410
Sequence: 1 QDLSSGAGCGEGYSRDATC.....ARATITRSGQTLKSWYNCP 1380

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Swissprot_39:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

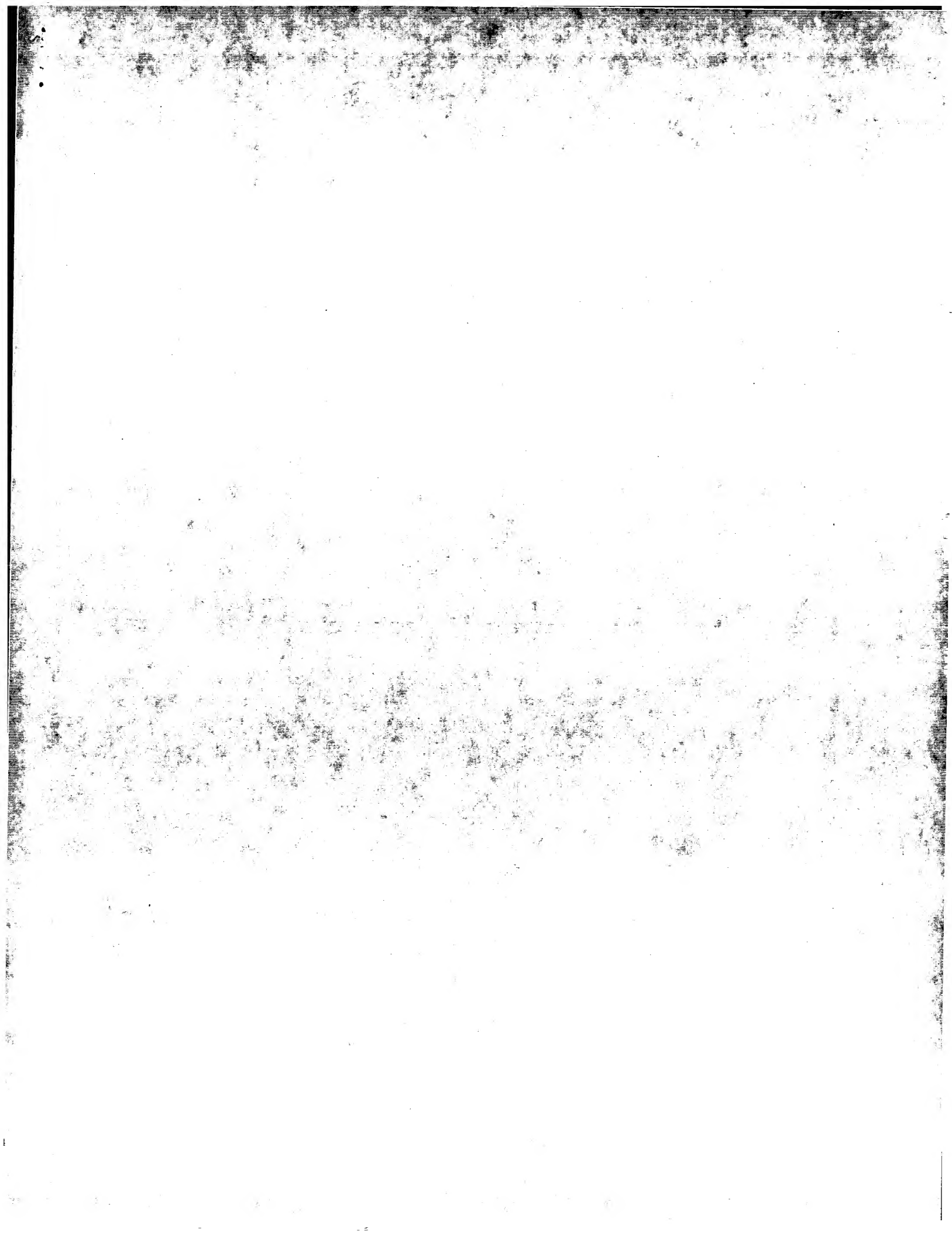
SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1188	16.0	5179	1 MUC2_HUMAN	Q02817 homo sapien
2	950	12.8	1664	1 SLPI_CLOVM	Q06852 clostridium
3	809	10.9	1367	1 AMYH_YEAST	P08640 saccharomyc
4	651	8.8	1255	1 MUC1_HUMAN	P15941 h mucin 1 p
5	559	7.5	875	1 FPI_MYTED	Q25460 mytilus edu
6	555.5	7.5	2700	1 ZAN_HUMAN	O9Y493 homo sapien
7	551	7.4	620	1 EXTN_TOBAC	P13983 nicotiana t
8	530.5	7.2	1162	1 TCNA_TRYCR	P19246 mus musculu
9	530.5	7.2	1162	1 CPN_DROME	P23253 trypanosoma
10	530.5	7.2	865	1 FPI_MYTCO	Q02910 drosophila
11	518.5	7.0	872	1 NFI_RAT	Q25434 mytilus cor
12	517.5	6.9	831	1 NFI_RAT	P16884 rattus norv
13	509.5	6.9	662	1 MUC1_XENLA	Q05049 xenopus lae
14	503.5	6.8	1970	1 RPB1_HUMAN	P24928 homo sapien
15	502	6.8	1020	1 NFI_HUMAN	P12036 homo sapien
16	498.5	6.7	1970	1 RPB1_MOUSE	P08775 mus musculu
17	493.5	6.7	467	1 RPB1_CRICR	P14141 cricetus
18	490	6.6	2142	1 BAT2_HUMAN	P48634 homo sapien
19	488.5	6.6	826	1 SSP_PLAYO	Q01443 plasmodium
20	475.5	6.4	267	1 EXTN_MAIZE	P14918 zea mays (m
21	471.5	6.4	5376	1 ZAN_MOUSE	O88799 mus musculu
22	468.5	6.3	634	1 VRP1_CANAL	P46593 candida alb
23	467	6.3	817	1 VRP1_YEAST	P37370 saccharomyc
24	454	6.1	797	1 VGEX_HAVER	P28968 equine herp
25	454	6.1	1161	1 YJ9P_YEAST	P47179 saccharomyc
26	453.5	6.1	1794	1 YAVI_SCHPO	Q10172 schizosacch
27	448.5	6.0	670	1 VG50_HSV1	Q00130 iclatulrid h
28	446	6.0	1229	1 NI21_HSV1	O9Y2N3 homo sapien
29	444.5	6.0	2774	1 MAPA_RAT	P34926 rattus norv
30	442.5	6.0	751	1 FPI_MYTCO	O27409 mytilus gal
31	439.5	5.9	1083	1 T2D3_HUMAN	O00268 homo sapien
32	436.5	5.9	1411	1 TCOR_HUMAN	Q13428 homo sapien
33	434.5	5.9	3256	1 KI67_HUMAN	P46013 homo sapien

34	433.5	5.9	3164	1 TEGU_HSV1	P10220 herpes simp
35	432.5	5.8	439	1 XP2_XENLA	P17437 xenopus lae
36	428	5.8	2517	1 NCR2_HUMAN	O9Y618 h nuclear r
37	427.5	5.8	1251	1 YOUT_CAEEL	O09350 caenorhabd
38	426	5.7	2715	1 TRX2_HUMAN	O9um66 homo sapien
39	425.5	5.7	1185	1 DRPL_HUMAN	P54259 homo sapien
40	424.5	5.7	1125	1 MAP4_MOUSE	P27546 mus musculu
41	424	5.7	307	1 SCS3_DROME	P02840 drosophila
42	423.5	5.7	3421	1 TEGU_HAVER	P28955 equine herp
43	422.5	5.7	1183	1 DRPL_RAT	P54258 rattus norv
44	419.5	5.7	2476	1 ZAN_PIG	O28983 sus scrofa
45	411.5	5.6	907	1 VGP3_EBV	P03200 Epstein-Bar

ALIGNMENTS

RESULT	1	ALIGNMENTS
ID	MUC2_HUMAN	STANDARD; PRT; 5179 AA.
AC	Q02817; Q14878;	
DT	01-JUN-1994 (Rel. 29, Created)	
DT	01-NOV-1997 (Rel. 35, Last sequence update)	
DT	20-AUG-2001 (Rel. 40, Last annotation update)	
DE	MUCIN 2 PRECURSOR (INTESTINAL MUCIN 2).	
GN	MUC2 OR SMUC.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.	
OX	NCBI_TaxID=9606;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	TISSUE=Intestine;	
RX	MEDLINE=94132002; PubMed=8300571;	
RA	Gum J.R., Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;	
RT	"Molecular cloning of human intestinal mucin (MUC2) cDNA.	
RT	Identification of the amino terminus and overall sequence similarity	
RT	to prepro-von Willebrand factor."	
RL	J. Biol. Chem. 269:2440-2446(1994).	
RN	[2]	
RP	SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.	
RC	TISSUE=Colon;	
RX	MEDLINE=93016075; PubMed=1400449;	
RA	Gum J.R., Jr., Hicks J.W., Toribara N.W., Rothe E.-M., Lagace R.E.,	
RA	Kim Y.S.;	
RT	The human MUC2 intestinal mucin has cysteine-rich subdomains located	
RT	both upstream and downstream of its central repetitive region."	
RL	J. Biol. Chem. 267:21375-21383(1992).	
RN	[3]	
RP	SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.	
RC	MEDLINE=91358717; PubMed=1885763;	
RX	Toribara N.W., Gum J.R., Jr., Culhane P.J., Lagace R.E., Hicks J.W.,	
RA	Petersen G.M., Kim Y.S.;	
RT	"MUC-2 human small intestinal mucin gene structure. Repeated arrays	
RT	and polymorphism."	
RL	J. Clin. Invest. 88:1005-1013(1991).	
CC	- FUNCTION: COATS THE EPITHELIA OF THE INTESTINES, AIRWAYS, AND	
CC	OTHER MUCUS MEMBRANE-CONTAINING ORGANS. THOUGHT TO PROVIDE A	
CC	PROTECTIVE, LUBRICATING BARRIER AGAINST PARTICLES AND INFECTIOUS	
CC	AGENTS AT MUCOSAL SURFACES.	
CC	- SUBUNIT: MULTIMERIC.	
CC	- SUBCELLULAR LOCATION: SECRETED.	
CC	- TISSUE SPECIFICITY: COLON, SMALL INTESTINE, COLONIC TUMORS,	
CC	Bronchus, Cervix and Gall Bladder.	
CC	- PTM: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR	
CC	INTERCHAIN DISULFIDE BONDS (BY SIMILARITY).	
CC	- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC AND	
CC	VARIABLE AMONG DIFFERENT ALLELES.	
CC	- SIMILARITY: THE N-TERMINAL DOMAIN SHOWS SOME SIMILARITY TO THAT	
CC	OF SLIKWORM HEMOCYTIN.	
CC	- SIMILARITY: CONTAINS 2 WFC DOMAINS.	
CC	- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTEINE KNOT-LIKE DOMAIN (CTCK).	



Db 2258 -----STEOPTSTPSGQSLTPMNSNSEVLTTSEPHVLSLSLSPDVQSSTTFNNLSESS 2312
QY 533 TKPEAPPTKEPAPPTPKPKAPPTTKEPAPPTKEPAPPT----- 572
Db 2313 TVE-----TPTSESVLSNSEPSTTEAPPTLSLDILSTTTNNLSQSSYSTVEDRSEISSE 2368
QY 573 -TKPAPPTAK-----EPAPPTKEPAPPTPKKLPTTPEKLAPPTKEPAPPTP 621
Db 2369 NSEKPT-SAPELVTSVTHVASSSPDVPTES--SEPDDLGSSTENIPKASSQITTSIP 2425
QY 622 EELAPPTPEEPPTPEEP-----APTTPKAAPNTKEPAPPT----- 660
Db 2426 TPTPTTASEEPPTKSTMSNPSTLSTSVNLSESSSTPESS-----KSPVSSSTEGISVVT 2481
QY 661 --PKPEAPT-----TPKEPAP-----TTPKEPAPPTPKGATPTTKEPAPPTPKP 704
Db 2482 EFSKVPESSTISSVLEBDLTKTTPSPILTEGTTASERTSEPLEDSLVASVNIHLTTSSSEN 2541
QY 705 APRELAPPTTKEPTSTTSOKPA-----PT-----TP--KGTAP- 735
Db 2542 VPKESESTTSSSESKPSOEPAIGLSTVAVPTSSVSLITASEIEAITSMTPEKQRTPI 2601
QY 736 -TTPK-----EPAPPTKEPAPPTPKGAPPTTKEPAPPTPKKAPKAPKELAPTTKGP 787
Db 2602 TTSKPKLVKSTSPSTVTSSEPSSESTKRTTVSTTVTTEETTTSESLILTAAPSKPI 2661
QY 788 -STTSOKAPPTPKETAPPTTKEPAPPTPKKAPPTPETPPTTSVSTPTTKEPTTIH 846
Db 2662 ESTSESEAPPTPAKTSERKPSNVSTSKRSTENETSTQSGLSSIMSSS----- 2715
QY 847 KSPDESTP--ELSAEPTPALE--NSPKEPGVPTTKPAATK-----PEMTTAKKTE 897
Db 2716 SEPEINAPVTVSSSEASTTLEENSTSSP---TSSEASVLSLPESTISEAVTVSS 2771
QY 898 RDLRTTPEPTTAAPKMTKATTTTEKTESKITATTT-----QVSTSTTQ 942
Db 2772 R-----APAEITMSSEHREISVTSSESEDEIPLSSTVSPNVTVASSISSEIILSVTS 2827
QY 943 DTPP-FKIT-----LKTTLAPKVTYTTTKTTTEIIMKPEETAKPKDRATNSKATTPK 996
Db 2828 SSTPRVRLTGGPDDLIVTVPSHGNRONITASSV---PSNSTPILPSES-LTTPQ 2883
QY 997 POKPTAPKPTSTKKPKTMPVRKPK--TTPTPKRMSTJMEPL-----PSTRIAEAM 1048
Db 2884 PPTTTTAKPATTSGKRPPSIQPPAEKMTTPAP-----PPPSNGCGEETNOEEOV 2937
QY 1049 LOTT 1053
Db 2938 TSTTT 2942

RESULT 14
151618
nucleolar phosphoprotein - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 24-Sep-1999
C:Accession: 151618: 5575757
J:Cell: C: McStey, B.
J:Cell: Sci 108, 3339-3347, 1995
A:Title: Identification and cDNA cloning of a Xenopus nucleolar phosphoprotein, xNopp180
A:Reference number: 151618; MUID:96019267
A:Accession: 151618
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-990 <CAI>
A:Cross-references: EMBL:X88927; NID:9895920; PIDN:CAA61368.1; PID:9895921
C:Genetics:
A:Gene: xNopp180
C:Superfamily: nucleolar-cytoplasm shuttle phosphoprotein
C:Keywords: phosphoprotein

Query Match 8.7%: Score 643.5; DB 2; Length 990;

Best Local Similarity 24.2%: Pred No. 1 le-19;
Matches 246; Conservative 148; Mismatches 424; Indels 199; Gaps 40;
QY 88 SPSSKAKAPPSSG-----ASQITKSTTKRSPKPNKKKT-----KVIESEETTER 133
Db 57 SPDAKKRPPANGLPKKAKSSSESSSEDEDEPPAKKRAQPAQAGKVVAVQPKKA 116
QY 134 HSYSENOESSSSSSSSSTTIK-IKSSYNSANAELOKLLKVKDNKKNT-----KK 186
Db 117 KSSSEDSDESEETTKKPPAKRPAQTPKVAAVKPTQKAKASSSESSSSSEDEASKK 176
QY 187 KPTKPP-----VDEAGSGDNGDFKVTPTPTSTQHNKSTSPKITTAAPINRPSLPN 243
Db 177 QPVITKVPQAVKAGLASNG-----KITADSSSEDSDESPAKKTATK-----TTP- 224
QY 244 SDTSKETSLVKNKETTVEKETTITTKQSTQDKERTTAKETOSIETKSADLAPTS- 301
Db 225 --TKPATAAKPOAKKTAGKSSREDSDEQKTAASKPKPDVSAVP---PTSVS 279
QY 302 --KVLAKPTPKAETTTKGALTTKPEPTPTPRE--PASTTPEPTTIKSAPTPKPE 357
Db 280 KKKTLSPGTAKAPESSDSDSDSEEOPAKKAIVPAKAAASAPKPLAKKATSTDSIS 339
QY 358 APTT-----TKSAPTTKEPAPTTTKEPAPTTKEPAPTTTKE-PAPTTYSAP 405
Db 340 DSSSEDEKSSVKGAKAP--KKAPA-----APDAKSTVPAAKKSAKPAKASSSD 390
QY 406 TTPKEPAPTTKPKAPPTTKEPAPTTKEPAPTTTKEPAPTTKEPAPTTTKEPAPTTAK 464
Db 391 SDSSSEETTTTPAKATTTAKSA-ATPTSTKTPNG-APTSTPAKPTPKSTAKKS 448
QY 465 KPAPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 524
Db 449 SSSSDSSSDSEETTTKPAKATTPAKASATPTSKPTNKAIPTSKKTAKCOTPTSNA 508
QY 525 PKEPSTTTKEPAPTTKPEAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 583
Db 509 KKDSSSDSSSDSEKTPA-----KRAKTTTPAKAATTTAKPA-----AKTTPAK 557
QY 584 PAPTTPKEPAPTTPKKLPTTPEKLAPTTPEKAPPTPELAPTTPEEPPTTPEAPPT 643
Db 558 PA-----AKSTPGKOVPTKKE-----SSSDSDSSSEDEKKS 590
QY 644 TPKAAPNTKPEAPPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 703
Db 591 SAKPAVKTT---PGKATSK---PVVASKPVP--AKRASSSDSDSSEETTKTKPLTKLS 643
QY 704 PAKELAPTTKEPTSTSD-----KAPTTTGTATTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 756
Db 644 PAVKTLPPKKAESSSDSDSEKKTAKPPAKSAPVNTKAPAKAONKASASCSDSDS 703
QY 757 TAPTTLEPAPTTPKKPAKELAPTTTGTSTSDKPAPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 816
Db 704 SSEEGRKSKQF-TGKSPAKATAP--PKKNPVAVNDKPSSSSDSDSGDEQ-----KPK 757
QY 817 KAPPTTPEPTTTEVSTPTTTKEPTTIHKSPPDSPELSAEPKALENSPKKPGVPT 876
Db 758 QAA-----AAKDVOKGAKKAPTKKAASSSSSE-----DS 787
QY 877 TTPPATPKPEMTTAKDKTTEREDLPTTPEPTTAAKMKETATTTTEKTESIITATTTQV 936
Db 788 SSEDVDSKAKNTNIVASNPV---TTKAVYAAKKESSSESDSEDEKQGGKNTSTIKI 843
QY 937 TSTTQDTPPEKITTLTKTTTLAPKVTYTK--KTITTEIMNKPEETA-KPKDRATNS--- 990
Db 844 ANST-----PKAAAECSSESSSEDEGKANGCTSGKRKRESIGNMEC 885
QY 991 KATTPKPKPTKAPKPKPTSTKKPKTMPVRKKTPTP--RKKTSTJMEPLNPTSLAE 1046
Db 886 EAVTPE-----NKKLKAASPTNFPVKNKELNTEPFRVREEDIEINP--RMAD 932

RESULT 15

QY	428	APPTKEPPTTPPKKEPAPTTKEBPAT--TPKEBPATPAKKAAPTTPKEBPATTP--PKEPAP	484
Db	2153	ALSSPAPSSRTTESPPSPGPTTTCGTGRTGTSRTTATATAPSKTRSTLLPSSPTSAPIITTV	2212
QY	485	TTTTKEPS-----	491
Db	2213	TTTCEPQACAMSEMWDI SYMPGPSSGDDPTYSNIRAGAGVCEBPLGECHRAQAQPVPL	2272
QY	492	-----PTPKKEBPATTTKAP	507
Db	2273	RELGOVVECSLDGVLICNRREQVGKFKMCFNEIRVFCFCNYGHPSPATPSTATPSSTP	2332
QY	508	TTTT--KEBPATTTKAPPTPKESPSTTTKEBPATTPPKKEBPATTTKKAAPTTPPKBPATTP	565
Db	2333	GTWIMLTKLTTTATTTTESTGSTATPSSITGPGAPHP--VSTTATPTVIVISSK---ATP	2386
QY	566	KEBPATTTKKA-----PTAPKEBA-----PTTPKEATPTPKKLTP-	602
Db	2387	FSSPBGATAPALPALKSTATTPTATSTALPSSSLGTTWTRLSTOTTPMATMTATAPSSTPE	2446
QY	603	-----TTPEKLAPTTPKEBPATTT-----PEELAPTTPEEP-	632
Db	2447	TVHTSVLTTATTTATGATGSVAATPSSPTGATTTKVPTTTTGTGVTPSSSPGATAPRPV	2506
QY	633	-----TPTPPEBPATTPKKAAPTPKKEBPAP--TPKKEBPATTP--TPKEBPATTP--TPKE--T	680
Db	2507	WISTTPTTSGSTVPSSIPGTHTPVLTITTPQVAVGSMAMPSSSTGSPSLIT	2566
QY	681	APTTPKCATATTLKEBPATTPPKKAPKELAPTT--TKEPS-----TTSOKAPATTT	729
Db	2667	TATTTATAGSTT--NPSSTPGITTPIPPELLTATTPATSSIVTPSSALGTHHPVNT	2624
QY	730	PKGT--APTTP-----KEBPATTPKEBPATTPKGATTP--TLK	763
Db	2625	TATTHGSLSPSSPHVRTAMTSATSGTLGTHHIEBPSGTSHMPAATGTTTSTPALS	2684
QY	764	EPAP-----TPPKKAPKELAPTTTKGTSTTSOKAPATTPKKEAPATTP--PKBP-----	810
Db	2685	SPHSPSKRTTESPPSP-----GTTTPGHTTATSKRTTATAPSKTRSTLLPSOPTSAPITTV	2740
QY	811	-----	810
Db	2741	VTTGCEPQACAMSEMWDI SYMPGPSSGDDPTYSNIRAGAGVCEBPLGECHATAPQPV	2800
QY	811	-----APTTPKKEBPATTPPTP	826
Db	2801	LGEIGOVVECSLDGVLICNRREQVGKFKMCFNEIRVFCFCNYGHPSPATPSTAMPSSST	2860
QY	827	PPTT---SEVSTPTTKEPTTIHKSDESPHLSAPPTKALENSKEGVEYTK-----	878
Db	2861	PGTWIMLTELTTTATTTAASGSTATP--STTP--GTAPPKVLTSPATTPATATSSATSSS	2917
QY	879	-----TPAPTK-----PEMTTAKDKTERTDLRTTPETT	907
Db	2918	SPKRTATILPVTISATATSTATSVTPPIPSLIGTGLPGLPOTITPAVMKTIHPSSTPETT	2977
QY	908	TAAPKMTKETATTTTETK-----TESKTATTTTOVTS--TTTQDTP-----	946
Db	2978	HTSVLTLTKATTRAATSSSTPSSIPGWTWILTELTAATTTAGTGTATPSSSTPGTWTWL	3037
QY	947	EKITTLKTTTLAPKVTTT--KTIITTEIMKKPEE-----TAKPKDRAT--NSKATTPKQO-	998
Db	3038	TELTATTTATTTASTGSTATLSSPTGWTWILPEBPTTAVTAAPRGSTVTAASSTOATGATGPHV	3099
QY	999	-----KPTKAPKAPKTPSTKKKPT--MPRVKPKTTPPKMT-----	1032
Db	3098	STTATPTPVTSKSKATPSSSPGATAPALPALSTATTTPTAHSFTAIPSSSLGTTWTLSTQT	3155
QY	1033	-----STMPELNPTSRILAEAMLOTT-----TRPNQTPNSKLVVNPKSEDAQ-	1074
Db	3158	TPPTAMSTAPSTSSRPENYHSTVLTATTTATGATGSVAATPSSSPGATHTTKVPTTTTGTG	3211
QY	1075	GABGEPHMLLRPHVMEPVETPDMQDLYPRVNOGIIINPMLSDENICNGKRPVGLTTL	1133

```

Db      3218  TATPSSSGTALTPPVWVSTTTPTTTP-----TSGSFVTPSSIGTHTARVLTITTTT 1184
Oy      1134  RNGTLVAFRGHFMWLPSPSPARRRTTEWGLPSPIDVYFTKCNCEGKT 1184
Db      3276  ATGSM-----ATPSSS---TQSGTPPSTLTATTTATGT 3309

RESULT 13
T34513
hypothetical protein ZK783.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34513
R:Favellio, A., Vaudin, M.
submitted to the EMBL Data Library, August 1994
A:Description: The sequence of C. elegans cosmid ZK783.
A:Reference number: Z21536
A:Accession: T34513
A:status: preliminary; translated from GR/EMBL/DBJ
A:molecule type: DNA
A:Residues: 1-3507 <FAV>
A:Cross-references: EMBL:U013646; PIDN:AC24418.1; GSPDB:GN00021; CESP:ZK783.1
A:Experimental source: strain Bristol N2; clone ZK783
C:Genetics:
A:Gene: CESP:ZK783.1
A:Map position: 3
A:introns: 14/1; 48/2; 84/1; 196/3; 303/1; 381/1; 586/1; 605/1; 1175/3; 1207/1; 1409/3504/1

Query Match      8.9%; Score 658.5; DB 2; Length 3507;
Best Local Similarity 22.6%; Pred. No. 8,7e+20;
Matches 286; Conservative 178; Mismatches 484; Indels 317; Gaps 51;

Oy      3  LSSCAGRCGEGYSRD-ATC-----NCDYNCOHMECCPDFKRVCTAELSCKGRCFESF 54
Db      1781  MGSCGCKCMAGYTGDDGATCIIKEEPRKSDKA----CTDEWSRLCELE--KRCQTVDE 1832
Oy      55  ERKREC-----DCAQCKKXDKCOPVESCACVHNHNTSPSSKAP 97
Db      1833  EEVPOGACLPGHHPINGTCQSLQISGLCAQKNDCKHAE--CIDH-----PDSHFCSC 1885
Oy      98  PSG-----ASQIKSTKRSKPPNKKTKKVESEIT 131
Db      1886  PDGFIDGMICDVDECNNAQMCDENTKCEMTGISFNCVLEGRKVKDEKCVDEKKOP 1945
Oy      132  EHSVSENGESSSSSSSSSSSTIWKIKSSKNSANRELQKKLVKXDKNKNPTKKKPPK 191
Db      1946  NREKIEIDENSSSSNSG-----OEKPTTK 1970
Oy      192  PRVYDEAGSLGNDGKVITTPDITSTQHNKVSPIITAKPINRPSLPPN---SDTSK 248
Db      1971  GIYVSTSS-----ATSSSETTAEPHVHTSISSTSTYKDMTSSKSPENVYMSSEEP 2019
Oy      249  ETSIVTWKETTIV-ETKETTITNKQSTDGKEKITSKETSQSIJEKTSANGLADP---TSKV 303
Db      2020  EVSTSSSKSTTASSETIVSSTPSSSSSEAPLINSPTATTEVITESSVKSTTPKESSSEI 2079
Oy      304  LAKPPTKAETTTKGPALTTPKPEPTTPPKPEPASTTPKEPPTTIS-APTTKREPAPTT 362
Db      2080  TVLKSSKSPVETSSVSKSPSPS--TTSQSVNISTVETSKSTVLSSEAEVSTSPTEVHT 2138
Oy      363  KSAPTTPKEAPPTTTPKEAPATTPPKREPAPTTTKSAPTTPKEAPPTPKKAPAT 422
Db      2139  -SSETKSLASASTGDTGNTSTPSTSSLASVKTSAPEGTSAAVAAYKLSSLSPDVSOPS 2197
Oy      423  TPK-EPAPTTPKPEPTTPKEAPTTKEP-APTTPEAPATPAKKAPATTPKEAPATPK 480
Db      2198  TKTFDATESSTVQASSETISSETSVKSTSEPSHNTKLSTISSNSSSVPTSPSPSTPVE 2257
Oy      481  EPAPTTKESPSTTP--KEPAPTTTKSAPTTTKREPAPTTTKSAP-----TPKESPSTT 532

```

Query Match 9.0%; Score 666.5; DB 2; Length 7962;
 Best Local Similarity 24.2%; Pred. No. 9.1e-20;
 Matches 268; Conservative 120; Mismatches 491; Indels 229; Gaps 50;

```

QY 74 PDYSCFAEHNHNPSPSSSKAPPPSGASOTIKSTYRS-----PKPPNKK 119
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6827 PEEEEFVE--EEVLPVKKRVKVPVAPVPEIKKKVKVVPKKEAPPAKVAPEVKKV 6884
QY 120 KTKVI--ESEEITEEHSVENOESSSSSSSSSTIWKSSKNSAARELOKKLVK 177
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6885 EKKRIIPKEEVLVPVYVEEPEEPISEE-----EIPPEEPIEEVEVAPRPV 6935
QY 178 DNKKNRKTKKPPRPVVDAGSLDN--GDFKVTTP-DTSTQHNKVNSTPKITTAKP- 233
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6936 EVIKKAVPEAPTPVKKVEAPPAKSKIIPEEKVVPVQKKAPPAVPEPKKVPKKV 6995
QY 234 -INRPSLRP--NSDTSKETSILVNKETVEK--ETTTTKQSTQCKE----- 278
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6996 LVPKKEAVPAKGRVLEEVSVAFQOEVVVKERLEVEAEVEEIPPEEPEHEVEYF 7055
QY 279 -----KTTSAKETOSIEKT-----SAKOLAPTSKVLAKPTPAETTKG 317
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7056 EEEGFHEVEEFIKLEQHRVEEHRVENVHVIYFEAEVEVEKPAAP-----KG 7107
QY 318 PALT-----TPKEPTPTTP--KEPASTPKPEPTPTTIKSAPTTKEP--APTITKSAPTTPK 370
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7108 PEISEKIIIPKKPPTKVPKPKKEPPAKVEVPKIVVEKVRVDEEPRVPTKVEVLPRK 7167
QY 371 EPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKSAPTTKAPPAPTTPK---PAPTTPK 426
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7168 EVVP-----EKKVNPVPAKKPEAP-----PKVPEAPKEVEVPKVPVPPKKKEV 7213
QY 427 PAPTKEPTPTTPKKEPAPTTKKEPAPTT-----KEPAPTAAPKAPPAPTTKEPAPTT 479
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7214 PPTKVEVPKAAVPEKKVPAIPKPESPPEVEEPEEESPAPK--KPVVPVVRKPEVP 7272
QY 480 KEAPPTTKESPPTTKEPAPTTTKSAPTTTKEPAPTTKSAPTTKEPSPPTTKEPAPT 539
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7273 KEVVPKPKVPAAP--PKKEPVPVK--VPEAPKEVPEKKVPV--PKKKPEVPPTK----- 7223
QY 540 TPKEPAPTPPKKAPPAPTTKEPAPTTTKEPAPTTTKKAPAPAKKEPAPTTKEPAPTTPK 599
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7324 VEPVPAVPAPEKVP-----EAIKPKESPPEVEEVEEVALEPAVEVEEPEAPAPQ 7379
QY 600 LT-----PTTPEKLAPTT---PEKAPPTTPEELAPTTPEEPTP--TTPPEEAPPTPKAAAP 650
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7380 VTPVPAKVPPEKAPAVVAKKPELPVVKVEVPEKVPVLPVPPK--PEAPAPKVP 7437
QY 651 NPPEKAP-----TTPKEP-----APTTPKEPAPTTKEPAPTTPKGTAFTTKEPAPTT 700
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7438 EVPKKEVVPPEKKAAPKPEVPKAPKVPV-----LEEKPAVP--VPERAESPPPEV 7490
QY 701 PKKPAKELAPTTTKEPSTTSDKAPATTPKGAPATTKEPAPTTTKEPAPTTPKGTAP 760
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7491 YEE--PEELAP---EEELAPPEEKVPVAEE-----EEVPPPAVPEEKIIPEKKVP- 7540
QY 761 TKEPAPTTPKKAPKELAPTTTKEPSTTSDKAPATTKEPAPTT-----PTTPKEPAPT 813
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7541 VAKKPAAPPKPEEPEKV--IEKPKLPRPPPPAPPAKEDVEKIFOLKAIIPKKVDE 7597
QY 814 TPKEP--APTPEPTPTTSEVS--PTTTPKEP-----TTHKSPDSTPELSAEP 862
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7598 NQVPEKVELTLPKVPGGKKVRLPERKPEKEVVLKSVLRKPEEPEEPEV--P- 7653
QY 863 KALENSPEKGVPTTKTPAATK---EMTTAKDKTTERDRTTPETTTAAAPKMTKETA 918
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7654 KMLE--KVKKPAVP--EPPEPKVEVEVPTTKRERKIPPEPKVPEIKAPIDLPAPDEK 7709
QY 919 TTTTEKTEKITAATTVQVSTTODTTPKITTLLKTTTLAPKVTTTKKITTTEINMK-- 976
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7710 PKPEA-----EVKTIKRPVEPEPTPIAAVTVPAVVGKAAE 7745

```

```

QY 977 ----PEEKAPKDRATNSKATTPPKOKPTKAPKKPTG--TKKKTPMPRVKPKTTPTP-- 1028
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7746 AKAPKEAPAKGPIKVPKKTPTSPIEARRKLRPGGGEKPPDEAPBTYOLKAVPLKVF 7805
QY 1029 ---RKMTSTMEELNPTSRIAEAMLQTTT 1053
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7806 KEIKDILTSESEFVGSSAIFECVLSPST 7833

```

RESULT 12

T45025
 mucin MUC5B, tracheobronchial [imported] - human (fragment)
 C:Species: Homo sapiens (man)
 C>Date: 21-Jan-2000 #sequence_rev1sion 21-Jan-2000 #text_change 21-Jul-2000
 C/Accession: T45025
 R:Dessey, J.L.; Guyonnet-Duperrat, V.; Porchet, N.; Aubert, J.P.; Laine, A.
 J. Biol. Chem. 272, 3168-3178, 1997
 A>Title: Human mucin gene MUC5B, the 10.7 kb large central exon encodes various alter
 A:Reference number: 722899; MUID:97166151
 A:Accession: T45025
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-3570 <DES>
 A:Cross-references: EMBL:722496; NID:91834502; PIDN:CAA96577.1; PID:91834503
 A:Experimental source: placenta
 A:Genetics:
 A:Gene: MUC5B

Query Match 8.9%; Score 663; DB 2; Length 3570;
 Best Local Similarity 22.1%; Pred. No. 5.8e-20;
 Matches 396; Conservative 119; Mismatches 578; Indels 698; Gaps 72;

```

QY 30 MECCPDFKRYCTALELCKGRCFSEFGRGCDCAQCKYDKCCPDYE--SCAEV--NHP 86
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1581 VECSLDFGLVC-----RNRE-----QVGKF--KMCFVNEIRFCNCVGHCP 1619
QY 87 TSPSSSKKAPPS-----GASOTIKSTKRSPKPKPKKKTKKVE 126
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1620 STATSTATPSSPTPGTWTLTGQTTAATTANTGTAIPSSPTGTAAPP-----KVLV 1673
QY 127 SEETIEHSVENOESS-----SSSSSSSSSTIWKIKSS---KNSAANR--- 168
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1674 SQATTPTATSSKATSSSPATATTLPLVLTSTARKSATSTPTPSSLTGTTGTSQNRPH 1733
QY 169 -----ELQKILKVDNKNKTKKAPTEKPP-----VUDE-----AGS 200
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1734 PMATMSTIHPSSTPTHTSTVLTTRATTTTRATSSMSSTSPGTTWLTLELTAAATTA 1793
QY 201 GLDNG-----DEKVTTPDTSTQHN-----KVSFPIITAKPI 234
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1794 ALPHGTPSSPGTWTLTLESTATVTVPGSATATSSIRATAGTLKVLTS--TATTPPYI 1852
QY 235 NPR-----PSLPPNSDTSKETSILVNKETVET-----KETTTNKQSTDG 276
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1853 SSRATPSSPGTATLPAKSTATTPTATSVTAIPSSSGTATRLSQTTTTPATMSTAT 1912
QY 277 KEKTTSAKETOSIEKTSAKDLAPTSKVLAKPKPKAKTTK-----GPA 319
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1913 PSSPTPTVHTSTVLTATTTTRGSAVATPSSPTGTAHTTKVPTTTTGTAAAPSSSPGA 1972
QY 320 LTPP-----KEPPTT-----TPKE-----PASTTPKEPPT 344
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1973 LTPPVWISTTTTPTPTGTVTPSSIPGTHNTATVLTITTTTVAATGSMATPSSSTQTSGR 2032
QY 345 TIKKSAPTPKKEPAPTTTSAAP--TPPKPAAPTTT--KEPAPTT---TPKEPAPTTKKEPA 397
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2033 PSLTTATTTTATGTTNPSSTPGTTPIPVLTATTAATAAASSTVPSASALGTTHTPVP 2092
QY 398 PTTTSA-----PTTP-----KEPAPTTPKKPAPTT---PKPP 427
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2093 PMTATTHGRSLRPSSPHIVPTAATWSATSGILGTHITPSTGTSHTPAAITGTTQPSPT 2152

```

Biochem. Biophys. Res. Commun. 165, 644-649, 1989
 A:Title: Sequence analysis of the 5' region of the human DF3 breast carcinoma-associated
 A:Reference number: A36735; MUID:90088473
 A:Accession: A36735
 A:Molecule type: mRNA
 A:Residues: 1-142, 'Q', 144-162, 'Q', 164-168 <ABE>
 A:Cross-references: EMBL:M31823; NID:g181542; PDB:AAA35757.1; PID:g181543
 R:Maizawa, Y.; Miyauchi, T.; Hamanoue, M.; Ando, S.; Yoshida, J.; Takao, S.; Shimazu, H.
 J. Biochem. 112, 609-615, 1992
 A:Title: A novel core protein as well as a polymorphic epithelial mucin carry peanut agglu
 A:Reference number: JX0235; MUID:93123189
 A:Accession: PX0066
 A:Molecule type: mRNA
 A:Residues: 996-1011, 'ES', 1014-1017, 1018-1032, 'T', 1034-1037, 1038-1057 <MAS>
 A:Experimental source: gastric carcinoma cell
 R:Zhang, L.; Baruch, A.; Elroy-Stein, O.; Keydar, I.; Wreschner, D.H.
 FEBS Lett. 356, 130-136, 1994
 A:Title: Tyrosine phosphorylation of the MUC1 breast cancer membrane proteins cytokine
 A:Reference number: S51026; MUID:95080414
 A:Contents: annotation
 A:Note: undetermined tyrosine residues in the carboxyl-terminal non-repetitive region an
 C:Comment: This protein is length polymorphic. Individuals may have between 21 and 125 c
 partial repeats. The repeat shown is defined by SmaI nuclease sites.
 C:Comment: Serine and threonine residues in the tandem repeat domain are extensively gly
 C:Comment: For an alternative splice form without a tandem repeat domain, see PIR:S48146
 C:Genetics:
 A:Gene: GDB:MUC1; PUM
 A:Cross-references: GDB:120705; OMIM:158340
 A:Map position: 1q21-1q23
 A:Introns: 20/1; 62/3; 1165/3; 1184/2; 1230/1; 1270/3; 1320/3
 C:Superfamily: polymorphic epithelial mucin
 C:Keywords: alternative splicing; duplication; glycoprotein; phosphoprotein; polymorphis
 E:1-1344/Product: mucin 1 precursor, splice form A #status predicted <PRAE>
 F:1-62/Region: mucin 1 amino-terminal non-repetitive
 F:1-23/Domin: signal sequence #link PREA #status predicted <SIGA>
 F:1-19-29-32/Domin: signal sequence #link PREA #status predicted <SIGB>
 F:1-19-29-32-1344/Product: mucin 1 precursor, splice form B #status predicted <PREB>
 F:1-19-29-212.1033-1344/Product: mucin 1 precursor, epithelial tumor antigen splice form
 F:138-1017/Region: 20-residue repeats (GSTAPAHGVSADTRPA)
 F:1143-1344/Region: mucin 1 carboxyl-terminal non-repetitive
 F:1245-1272/Domin: transmembrane #status predicted <TRM>
 F:1046,1064,1118,1144,1232/Binding site: carboxylate (asn) (covalent) #status predicted
 F:1213/Binding site: phosphate (tyr) (covalent) #status predicted

Query Match 9.1%; Score 672; DB 1; Length 1344;
 Best Local Similarity 27.8%; Pred. No. 9.6e-21;
 Matches 310; Conservative 78; Mismatches 472; Indels 256; Gaps 56;

DB 252 LVNKKETVETKETTNNK--QTSDGKKTTSKKTOSIEKTSKAD-LAPTSKVLAKPT 308
 15 LVVLTATAPKATVTVGSGHASTPGGEKETATGKSSVSPSEKNAVMTSSVLSHS 74

DB 309 P-RAETTTGKP--ALTTPKEP-----TPTTPKEPASTTP-----KE 341
 75 PGSGSSTTGGQDVTLPAPAEPAAGSAAATGQDVTSVPTPALGSTTPRAHDVTSAPDNK 134

DB 342 PPTFT-----TKSAPTPPKKEPAPTT-----KSAPTPPKKEPAPTTKEP-----A 381
 135 PARGSTAPPAHGVTSAPDT--RPAFGSTAPPAHGVTSAPDT--RPAFGSTAPPAHGVTS 190

DB 382 PTPPKKEPAPTTKEPAPTTTTSAPTPPKKEPAPTT-----KKPAPTTKEP----- 427
 191 PDI--RPAFGSTAPPAHGV--SAPDTRPARGSTAPPAHGVTSAPDTRPARGSTAPPAHGV 247

DB 428 --APTTPKEPTPTT-----KRPAPTTKEP-----APTTPKEPAPTPK----- 464
 248 TSAPDTRPARGSTAPPAHGVTSAPDTRPARGSTAPPAHGVTSAPDTRPARGSTAPPAHGV 307

DB 465 KRAPTPPKKEPAPTPK-----EPAPTTTKEPSPTTPPKKEPAPTT-----TSAPTTTKEPAPTT 517
 308 TSAPDTRPARGSTAPPAHGVTSAPDTRPARGSTAPPAHGVTSAPDTRPARGSTAPPAHGV 367

DB 518 TKSAPTPPKKEPSPTTPKEP-----APTTPKEPAPTPK-----KRAPTPPKKEPAPTPK- 566
 368 T-SAPDT--RPAFGSTAPPAHGVTSAPDTRPARGSTAPPAHGVTSAPDTRPARGSTAPPA 424

DB 567 ---EPAPTTTTPKRAPYAPKEPAPTPPKETAPTPPKLPTTPPKLAPTTPEKAPPTPE- 622
 425 HGVTSAPDTRPARGSTAPPAHGVTSAPDTRPARGSTAPPAHGVTSAPDTRPARGSTAPPA 484

DB 623 ---ELAPTTPEETPTT-----EERAP--TTPKA-----AAPNTPKKEPAPTPK- 662
 485 HGVTSAPDTRPARGSTAPPAHGVTSAPDTRPARGSTAPPAHGVTSAPDTRPARGSTAPPA 544

DB 663 ---EPAPTTPKKEPAPTPPKETAPTPPKGAPTTLKEPAPTT-----KKPAPKEL 709
 545 HGVTSAPDTRPARGSTAP-----PAHGVTSAPDTRPARGSTAPPAHGVTSAPDTRPARGST 600

DB 710 APTTKEPTSTSDKAP--TPPKGAPTPPKKEPAPTTPEKAPPTPKG-----TAPPTLK 763
 601 AP-PAHGVTSAPDTRPARGSTAPPAHGVTS-----APDTRPARGSTAPPAHGVTSAPDTRP 655

DB 764 EPAPTP-----KKPAPKELAPTTKGPSTTSDKAP--TTPK-----ETAPTP 807
 656 APGSTAPPAHGVTSAPDTRPARGSTAP--PAHGVTSAPDTRPARGSTAPPAHGVTSAPDTR 714

DB 808 KEPAPTTP-----KKPAPTTPEPTTSEVSTPTTKEPT-----IHKSPD 850
 715 PARGSTAPPAHGVTSAPDTRPARGS--TAPPAHGVTSAPDTRPARGSTAPPAHGVTSAPD 772

DB 851 ESTPELSAETPKALENSKEPEVPPTTKITAA-----TKEMTTAKD--TTERD 899
 773 TRPARGSTAPPAHGVTSAPDTRPARGSTAPPAHGVTSAPDTRPARGSTAPPAHGVTSAPD 832

DB 900 LPTTPET-----TAPKMTKETATTTETKTTATTTGTTTODTTP--FKIT 950
 833 TRAPGSTAPPAHGVTSAP-----DTRPAPSTAPPAHGVTSAPDTRPARGSTAPPAHGV 888

DB 951 TLKTTTLAPVYTTTKTTT---EIMNKPETAKPKRAATNSKATTPKPKAPKAPKPT 1008
 889 SAPDTRPARGSTAPPAHGVTSAPDTRPARGSTAPPAHGVTSAPDTRAP--GSTAPPAH 946

DB 1009 STKKPKTMPVRKPKTTPTRKMTSTPELNP-----TSRLAEMLQT--TTRP--NOTPN 1060
 947 VTSAPDTRP--APGSTAPPAHGVTSAPDTRPARGSTAPPAHGVTSAPDTRPARGSTAP 1003

DB 1061 SKLVEVNPKESEDAGAGETPRHMLLRPHVPEVTPDMVILPRVNOGIIINPMLSDETN 1120
 1004 AHGVTSAPDTRPARGSTAPPAH-----GVTSAPDNKRALGSTA--PPVHNVT 1049

DB 1121 ICNCKPVDGLTTLNGLVAFRGHYFMMLSPFSPS 1156
 1050 ASGSASGASTLVHNGTSARATTPPAKSTPFSLPS 1085

RESULT 11
 I38346
 elastic titin - human (fragment)
 C:Species: Homo sapiens (man)
 C:Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 21-Jul-2000
 C:Accession: I38346
 R:Labelit, S.; Kolmerer, B.
 Science 270, 293-296, 1995
 A:Title: titins: giant proteins in charge of muscle ultrastructure and elasticity.
 A:Reference number: A57430; MUID:96026330
 A:Accession: I38346
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: mRNA
 A:Residues: 1-7962 <RES>
 A:Cross-references: EMBL:X00569; NID:g1017426; PDB:CAA62189.1; PID:g1017427
 C:Genetics:
 A:Gene: GDB:TTN
 A:Cross-references: GDB:127867; OMIM:188840
 A:Map position: 2q31-2q31


```

OY 49 RCFESFEGRECDCAQCKKDYDCPDYESFCAEVHNPTSPSSKAPPPGASGOTIKST 108
DB 34 RDKFKVHKIKKNTCTSC-----KCPD-----APSN-----PDVSTIIS- 70
OY 109 TKRSPKPNKKTKKVISEETIEHSYSENOESSSSSSSTTIKISSKSNANR 168
DB 71 -----LNDVNDIGPSGDSNPTGSSWFOIEAIVGOGATVSEHNIDSSV 114
OY 169 ELQKKLVKKNKKNKTKPPKPPVNDGASGLDNGDFKVTPTDSTTQ---HNKSTS 225
DB 115 EVEKKV-----TTSIDASTTNAPTTGKOST 140
OY 226 PKITTA-KPIMRPSLPNSDT-----SKETSLVKNKETTVEKETTNNKOTSDGK 277
DB 141 PELTIGVIVINSKESVYTDMSSTREFTLLSPETLVSSTDSST-EDTSPDNT 199
OY 278 EKTTSAKETOSIEKTSADLAPTSKVLAKPTPKAETTKGALTPKEPPTTKPEAST 337
DB 200 TEIASPMETINTTATTSVSPSTIAS---EDETVTAIAEST-----TVIAEVST 250
OY 338 TKREPTPTTIKSAPTTKREPAPTTKKAPPTTKREPAPTTKREPAPTTKREPA 397
DB 251 TTEPTPTT-----AESITTKSTT---KAPATTEEPPTTEE---VTTEAET 292
OY 398 PTTKSAPTTKREPAPTT-----KKPAPTTKREPAPTTKEP---TPTTKREPAPTTKEPAP 452
DB 293 STTSSSESTEK-----PTTPLIDNKIAPATGK---PETTHFPTVGTTPN---FDIATETPR 344
OY 453 TTKREPAPTKAPKAPPTTKREPAPTTKREPAPTTKRE-----PSPTTKREPAPTTKSAP 507
DB 345 VANSSEDMKTSKTAAMETTOOTTEVT-DGPEKETTKVNSIEITTVPLVETISTSTAS 403
OY 508 -----TTTKREAPT-----TTSAPTPTKREPSPTTKE---PAPT-----TPK--- 542
DB 404 KESDGEFTITLAKLVYTAADSSTASATVKEPNEETTKSHVAKPTKGTVKVPTELIS 463
OY 543 --EPAPTTPKPPAP-----TPKE-----TPKE-----PAPTTP 565
DB 464 FDEPTTEIT-KAPHGKLEKKTGYHFLSDNFANYSKAKENDDNHLDYNIHRAKERTPT 522
OY 566 KEPAP-----TTTKKAPTA---KEPAAPT---PKEAPPTPKLPTPEKLA---PTTP 613
DB 523 EESSTEEVETTEEPANGTPTTENPTTEOPTSTAESTYALAPPTTQGVTEBPTTA 582
OY 614 EKPAPTPEELATPEEPTPTTPEPAPTTPKAAPNPKREPAPTTKREPAPTTKREPA 673
DB 583 EKSTAGQ-----KPTTQGESVST---EKSTTKKA---STTEE---PTTDEPTTPT---ES 627
OY 674 PTTPKETAAPTTPKGIAPTTLKREPAPTTPKAPKELAPPTTKPEPTSTSDKRAP-----TT 729
DB 628 STTGKATPELSTSTSEETTTTELKITE-----GSTTEBPTTALFAASIGITTT 679
OY 730 PKGTAPTTKREPAPTTKREPAPTTPKGTAPTTLKREPAPTT----- 770
DB 680 DEETSTSTTTEITSTK-IVTESALITQTSVSVESSTPQLOLBERMKAIVNKEKHNL 737
OY 771 ----KKRPAKELAPTTKGP-----TSTTSKRAPPTPKETAPPTTKEP 810
DB 738 VLKEXKRLKRESSTGSDSEPTTVAAENIDEVTTTEKKKVVQITPTTEKSTIOEE 797
OY 811 APPTPKKAPPTPE---TPPTTSE-VSPPTTKEPTTTHKSPDESTPELSAPPTKALEN 867
DB 798 TTTTTEKTTSTKTTTEKPTTSASATETTTSEPTST---TEST----- 838
OY 868 SPKRGVPTTKPAATKPEMTJADKTERLARTPTETTTA----- 909
DB 839 -----TVODSSATTEESSIAETTTTSAE---TSETTSSAATTTGSPENTALQS 887
OY 910 -----APKMTKETATTTTEKTTESKITAATTO---VTST---TODT 944
DB 888 SSOQSEENESSAEMKPGARDPVRPKKHKTTPVAPETTSVAASSTTTTEITTEKSTTLET 947
OY 945 TPKFITTLTKTTLAPKV---TTKTTITTTTEIMKPEETAKPKDRANINSKATTPKPO--- 998

```

```

DB 948 TPTEATTLNEVTGPAETVGAPEVDETINTLELLSK-----INTTOISQPKPTDIS 997
OY 999 -----KPKAKKKPTSTKPKPTMVRKPKPTTPPKMTSTMPELN----- 1039
DB 998 KTDALSSLSLGLSGFTKAPMARTI-----HTTDAFVATVATEASLNDGSDK 1044
OY 1040 -----PSTRIAEA 1047
DB 1045 KIIDEAQPTDEIRRA 1059

```

RESULT 10

A35175

mucin 1 precursor, repetitive splice form A [validated] - human

N:Alternate names: breast carcinoma-associated df3 antigen; core protein K939; epistatin; mucin; mucin 1 precursor; epithelial mucin (PEM)

N:Contains: mucin 1 precursor; epithelial tumor antigen splice form; mucin 1 precursor

C:Species: Homo sapiens (man)

C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 02-Jun-2000

R:Rigdenberg, M.J.L.; Vos, H.L.; Gemmissen, A.M.C.; Hilkens, J.

J. Biol. Chem. 265, 5573-5578, 1990

A:Title: Epistatin, a carcinoma-associated mucin, is generated by a polymorphic gene

A:Reference number: A35175; MUID:90202794

A:Accession: A35175

A:Molecule type: mRNA

A:Residues: 1-952,1033-1344 <LIG1>

A:Cross-references: GB:M32738; GB:J05288; NID:g182121; PIDN:AAA35804.1; PID:g182124;

A:Experimental source: Splice form A

A:Note: Genbank entries HUMEPIS1 and HUMEPIS12 present only the amino and carboxyl

A:Accession: B35175

A:Molecule type: mRNA

A:Residues: 1-19,29-952,1033-1344 <LIG2>

A:Cross-references: GB:M32738; GB:J05288; NID:g182126; PIDN:AAA35806.1; PID:g182129;

A:Experimental source: Splice form B

A:Note: Genbank entries HUMEPIS1 and HUMEPIS12 present only the amino and carboxyl

R:Gendler, S.J.; Lancaster, C.A.; Taylor-Papadimitriou, J.; Duhig, T.; Peat, N.; Butc

J. Biol. Chem. 265, 15286-15293, 1990

A:Title: Molecular cloning and expression of human tumor-associated polymorphic epith

A:Reference number: A35886; MUID:90368715

A:Accession: A35886

A:Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-19,29-992,1033-1344 <GEN>

A:Cross-references: GB:J05581; NID:g188869; PIDN:AA59876.1; PID:g188870

A:Note: Genbank entry HUMMUCAB includes one copy of the tandemly repeated sequence

R:Ilan, M.S.; Batra, S.K.; Qi, W.N.; Metzgar, R.S.; Hollingsworth, M.A.

J. Biol. Chem. 265, 15294-15299, 1990

A:Title: Cloning and sequencing of a human pancreatic tumor mucin cDNA.

A:Reference number: A35887; MUID:90368716

A:Accession: A35887

A:Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-19,29-1109, 'S', 1111-1339, 'A', 1341-1344 <LAN>

A:Cross-references: GB:J05582; NID:g189598; PIDN:AAA60019.1; PID:g189599

A:Note: Genbank entry HUMPANMU contains four fewer copies of the tandemly repeated se

R:Wreschner, D.H.; Hareuveni, M.; Tesarfay, I.; Smorodinsky, N.; Horev, J.; Zaretsky,

Eur. J. Biochem. 189, 463-473, 1990

A:Title: Human epithelial tumor antigen cDNA sequences. Differential splicing may gen

A:Reference number: S10571; MUID:90276413

A:Accession: S10572

A:Molecule type: mRNA

A:Residues: 1-19,29-1155, 'P', 1157-175, 'P', 177-182, 'A', 184-212, 1033-1344 <WRE>

A:Cross-references: EMBL:X52229; NID:g937053

R:Wreschner, D.H.

Submitted to the EMBL Data Library, March 1990

A:Reference number: S40293

A:Accession: S40293

A:Molecule type: mRNA

A:Residues: 1-19,29-1155, 'P', 1157-175, 'P', 177-182, 'A', 184-212, 1033-1037, 'A', 1039-1344 <

A:Cross-references: EMBL:X52229; NID:g937053; PIDN:CAA36478.1; PID:g937054

R:Abe, M.; Siddiqui, J.; Kufe, D.


```

QY 739 -KEBAPTTKPEAPTTKPGATTLKEBAPTTKPKPAKELAPTTTKGPTSTSDKBPAP 797
DB 1001 MSSPPPEVKSPPPAPVSSPPPPVKSPPPAPVSSPP---PVPKSPPPAPVSSPPPPV 1057
QY 798 -TPKETAATTKPEBAPTTKPKKAPTTTETPPPTTSEVSTPTTKKEPTTIKSPDESP 856
DB 1058 KSPPPAPVSSPPPPVKSPPPAPVSS---SPPPVKSPPPAPVSSPPPPVKSPPPAPV 1115
QY 857 SAEPTPKALENSPKKPGVPTTKTPAATKEMTTAKDKTTERDRTTPEPT 906
DB 1116 SPPAPV-----VKPPSLP-PPAPVSPPPVTPAPPKKEGSLPPAPAS 1158

RESULT 8
high molecular mass nuclear antigen - chicken (fragment)
C:Species: Gallus gallus (chicken)
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T18535
R:Shimada, K.; Harata, M.; Mizuno, S.
J. Cell Sci. 110, 3031-3041, 1997
A:Title: A nuclear matrix-associated high molecular mass nuclear antigen, HMNA, of chick
A:Reference number: 218955; MUID:9803440
A:Accession: T18535
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1151 <SH1>
A:Cross-references: EMBL:D88440; NID:d1177138; PTD:d1025045; PTDN:BAA24137.1

Query Match 9.4%; Score 695; DB 2; Length 1151;
Best Local Similarity 25.3%; Pred. No. 9.4e-22;
Matches 304; Conservative 108; Mismatches 466; Indels 324; Gaps 59;

QY 81 AEVHNFTSPSSSK-----APPSASQTIKSTIRSPKPKPKKTKKVKIESEETEH 134
DB 39 AEIH-PAHQPPKWPVIGGAPPEPT-----PSKPLDGDAAAPKASALITSP 90
QY 135 SVSENESSSSSSSSSTIWKIKSSKNSAANRELQKLVKDNKNKTKKKPKPPPV 194
DB 91 PASPSPPDPKAPSGCAEAGTPPSQGPAG-----TPP-- 126
QY 195 VDEAGSLDNGDFKVTTPDTSTOHNVKSTS--PKITTAKPINRPSLPSNSTSKETSL 252
DB 127 -SGAAGAKRGDGTAPSGTKSGADGKPAQDVPAKATTA-AEAPR-----ASAASP 176
QY 253 TVNKETTVETKETTINK--QTSDGKEKTSKETSIEKTSKIDLPATSKVLAKPTPK 310
DB 177 TVKKAATATTAASASAPKATDAATAATTA---SOSAPKATV-EVKPAAAVAKENAKA 232
QY 311 AETTTGPAITTPKEPTTP--TPKREASTPKEP-----TPTTIKSAPTTKEPAPT-- 361
DB 233 VTAAAPAKATATAKAPVATSPITPCSSABAKDLTAASPTASKA--TAEAKVPATASLM 290
QY 362 -----TKAPTTKEPAPTITTK-----EPAPTKEPAPT 391
DB 291 ATKVTAAPKAPSPSVKATITDKAVATAAPKAGPDVKAIVANCAEAKPAPPPPOOLPK 350
QY 392 TTKEPAPTITTKSAPTT-----PKEP-----PTTKKAPPTKEPAPTTPKEPTTP 438
DB 351 AAAAAATGELKATPRPHGSPRANSHVTVPNVPRAAATVP--TAGAVPKASTGT 408
QY 439 T---PREPAPTKEPAPTTPKEPAPTAPKPKAPATTTKEPAPTTPKEP-----APTTPKEP 491
DB 409 TPAAAPQOVPV---KAAPVPPSPQOAVPPRAATAA---APVTPQOVTKAATVATNATP 461
QY 492 PTPKPEAPTTTKSAPTTKEPAP-----TTTKSAPTTKEPSPPTTKEP-----APTTPKEP 544
DB 462 PPOGTPRAATITTTATPTPPOGPIKAKTDAAAPRAVPAKASDGAATPGVPMATDPOKP 521
QY 545 APTPKKAPPTTPKEPAP-----TPKREAPT--TTKKAPATPAK-----EPAPT 587

```

```

DB 522 PPTQSVASVATEPKPOPRAPAPPSNEATPAVPSPNLKSPLPTIPKPVLMALTPOPV 581
QY 588 TPK---ETAPTTKELKLTPTTEKLAFTPEKPAFTTPEEL-----PTTP----- 629
DB 582 TAQWVTLAATKSPPIPKASPK-ALMTTPPPGGLRALAAKLLGLSPVASAMHAK 640
QY 630 -EEPTPTP-----EEPAFTTP-----KAAPTPKE-----PAFTTPKEP 664
DB 641 VTPEPLASVPMASASASISGLDAAARVALATNMAAPKAEAGNGCTLMAAPGAANTOM 700
QY 665 APTPKPEAPTTKETAAPT-----TPKTAFT----- 691
DB 701 APTGAAGAQTAPMGAATHVSPMAGAGATQMSPTGAANTHMSPIGAGATQMSPMGAAN 760
QY 692 TLKEPAPTTPKPKPAKELAPTTKEP-----TSTTSKPAFTTPKG--TAFTTPKEP 741
DB 761 TOMSPMGATTTQMSPMGAATQOSPMAAATQVATISAGTITMOVSFPGAATPPQTSVG 820
QY 742 APTPKPEAPTTKPGTAPTTKEP--APTTPKPKPAKELAPTTTKGP-----TSTTSKP- 794
DB 821 AATTP-QPSM-----GAATITMSPMGAATTPQ---PSPMGAVTTQPPMAATNTTQPPM 872
QY 795 APTTPKEP--APTTPKEP--APTTPKPK-APTTPETTP-----PTSEVSTPTTK 840
DB 873 AASTPQSTPMGAATTTQSPMGATTTQSPMGASTPQAPPAVAGSPPTPPPIPSPTAQT 932
QY 841 EPTTIHKSPDESTEP--SAEPTKALENSPKKPGVPTTKTPAATKPEMTTAKDKTER 898
DB 933 SPQMSKSPDPDPAPSAQAOTSPAHVANAASQV-TAVSPA---PIGVTEASPSDGA 988
QY 899 DLRTPETTTAAPKMT-KEPATTEKTESKITATTTQVSTTTQDTPPFKITLLKTTTL 957
DB 989 RLSPPTAATGPKASPAATADVTEAATD--VTAATATVPA-----EAAP-----TK 1033
QY 958 ADKVTITTKTTTITTIMKPEETAAPKRAIRNSKATTTKPKQPIKAPKPKPTSTKKPKTMP 1017
DB 1034 AKRSSSSSSSSSSSSSSSSSSSSSSSDSSSSSSSESNPSPA-----P 1078
QY 1018 RVRKKTTPTPPKMTSTMPLELNPTRIAEAMLOTTTRPNQTPNSKLVEYNKPSSEDAGAE 1077
DB 1079 AVGDGQOQMTPAQASVPP-----VTEANQ-----EAAAAAAGAAE 1117
QY 1078 GE 1079
DB 1118 RE 1119

RESULT 9
T25697
hypothetical protein F16F9.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T25697
R:Fulton, B.
submitted to the EMBL Data Library, August 1996
A:Description: The sequence of C. elegans cosmid F16F9.
A:Reference number: 220071
A:Accession: T25697
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1229 <F16F9>
A:Cross-references: EMBL:U67956; PTDN:AA07691.1; GSPDB:GN00028; CESP:F16F9.2
A:Experimental source: strain Bristol N2; clone F16F9
C:Genetics:
A:Gene: CESP:F16F9.2
A:Map position: X
A:Introns: 35/1; 361/1; 482/2; 517/1; 971/1; 1021/2; 1179/3

Query Match 9.1%; Score 676.5; DB 2; Length 1229;
Best Local Similarity 27.2%; Pred. No. 5.7e-21;
Matches 330; Conservative 95; Mismatches 385; Indels 405; Gaps 56;

```

Query Match 10.9%; Score 809; DB 1; Length 1367;
Best Local Similarity 27.5%; Pred. No. 2,3e-26;
Matches 328; Conservative 112; Mismatches 526; Indels 226; Gaps 53;

```
QY 61 DCDACCKKYYD---CCPEYSEFCAEVHNPISPSSKKAAPPSSGASQTIKSTTKRSPKPPN 117
DB 174 DLTGTCNNYDNOGHQSOTDFEFGYWNIDCDNNCGGTSTSTSSSES---STT----- 223
QY 118 KKKTKVYESEITEEHSEVENOSSSSSSSSSSSTLWKSSKSNANRELQKLLKLV 177
DB 224 ---TSSTSESTTSSSESTSSSTSSSESTSSST---APAPPTTSCREKPTP 274
QY 178 DNKNNRTKKATPKPPVVDDEAGSLDNGDFKVTTPDTSTQHNKVTSPKITTAKPIPR 237
DB 275 PTTTSCREKPT--PPIHD-----TTPCT-----KKTTTSCCTKTTTTPV 314
QY 238 PSLPPNSDTSEKSTLYNKKETVETKETTNNKOTSDGKEKTTSAKTOIEKTSKDL 297
DB 315 PT--PSSSTTSSS-----APVPTSSSTT-----ESSSAPVTSSTSSSS--- 353
QY 298 APISKVLAKPTPKAETTTGPAALTTPKEPTTPPKEPASTTPKEPTTISKAPTPPKPE 357
DB 354 APV-----PTPSSSTESSAPVTSSTTSSSAPVTSSTESSSAPVPTPSSSTESS 407
QY 358 APPT-----TKSAP-----TTPKEPAPTTPKEPAPTTPKEPAP-----TTPKEPAPT 404
DB 408 APVTSSTESSSAPVTSSTESSSAPVTS-----TTESSAPVTSSTESSSAPVPTPSS 463
QY 405 PTPPKEPAPTTPPKPAPTTPKEPAPTTPKEPTP-----TTPKEPAPTTPKEPAPTTP 461
DB 464 STTESSSAPVT-----STTESSSAPV-----PTPSSSTESSSAPV-----STTESSSAP 512
QY 462 APKKPAPTTPKEPAPTTPPKPAPTTPKEPAPTTPKEPAPTTPKSAPTTTPKSA 521
DB 513 -----PTPSSSTESSSAPVTPS-----STTESSSAPVTPS-----TTESSSAPVTPSS 559
QY 522 PTPPKESPPTTPKEPAPTTPKEPAPT--TPKKPAPTTPKEPAPTTPKEPAPTTPKKA 580
DB 560 STTESSSTPTSS-----TTESSSAPVPTPSSSTESSSAPVPT-----PSSSTESSSAPA 611
QY 581 PKKPAPTTPKEPAPTTPPKKILPTTPPEKLAPTTPPEKAPPTPEELAPT--TPKEPTTPPE 639
DB 612 PTPSSSTESSSAPVT-----STTESSSAP--VPTPSSSTESSSAPVPTPSSSTESSSA 666
QY 640 PAV-----TTPKAAAPNTPKPAPTTPKEPAPTTPKEPAPTTPKEAP--TPKGAAPT 693
DB 667 PVTPESSSTESSSAPVT-----STTESSSAPVT-----STTESSSAPVPTPSSSTESS 718
QY 694 KEAPAPTTPKPAKELAPTTTKEPTSTSDKPAPTTPKGAPTTPKEPAP--TTPKEPA 750
DB 719 SAVPTPSSSTESSSAPVPTPS--SSTTESSSAPVTSSTTESSSAPVPTPSSSTESSSA 777
QY 751 PT--TPKGAAPTTPKPAKELAPTTTPKPAKELAPTTTGGPITSTSDKAPAPTTPKETAPT 809
DB 778 PVTPESSSTESSSAPVPTPSSSTESSSAPVPTPSSSSNITSSAPSTPSSSTESSSSV 837
QY 810 PAV-----TTPKPAPTTPPEPTTSEVSTPT-----TTPKEPTIHKSPDESPTELSAE 859
DB 838 PVTPESSSTESSSAPVSSSTESSSAPVPTPSSSSNITSSAPST---PSSSTESSST 894
QY 860 PTPKALENPKKEGVPPTTKPAATPEKMTTAKOKTTERDLKTPPETTTAAPKMKETAT 919
DB 895 GT--TVPSSSKTPGSGOTETSVSTETETIVPKTTTSVTETSTTTTTCSTGNSAGE 953
QY 920 TTEKTESKITAT--TTOVSTTTODTTPPKITLTKTTLPKVT---TKKTTT--EI 973
DB 954 TTSKGSPTVTTTPTTTTSTTTTSTTTTTCSTGNSAGETTSKGSKTTTTPVPC 1013
QY 974 MKKPEETAKPKDRATNSAATTPKPOK-----TKAPKKTSTKKKPKT 1016
DB 1014 STSPETA-----SESTTSTPTPTVTIVSTTVTETSTSKGKGELTTTTFVKN 1065
```

```
QY 1017 PVRKPKPTTPPKKMTSTMBELNPTSRIAEAMLOTTTRPNQTPNSKLVENPKSEDA 1076
DB 1066 PTTVLTITIAPTP--SVTATVNFPTT--ITTVCSGT-----NSAGETTSKC 1109
QY 1077 EGEPTPHLLRPHFMFEVPPDMOYLPRVRNOGIIINPMILSDETNICONGPVDGLT 1136
DB 1110 SPKTVTTTTPCSTGTEYETATLTVTAVTTTVTSTESGTNSA--GKTTTGTTK 1168
QY 1137 TLVAFRGHYFWMLSPPSPSPARRITVEWIGIPSPIDVPTFC-----NCSCKT 1184
DB 1169 TT-----YVTLAPASAVTPATN-----AVPTTITT--TECSAATNAGET 1207
```

RESULT 7

extensin-like protein - maize
C:Species: Zea mays (maize)
C:Date: 05-Mar-1995 #sequence_revision 12-May-1995 #text_change 29-Oct-1999
R:Accession: S49915
R:Rubinstein, A.L.; Broadwater, A.H.; Lowrey, K.; Bedinger, P.A.
submitted to the EMBL Data Library, June 1994
A:Description: Pez genes: pollen-specific genes with extensin-like domains.
A:Accession: S49915
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-1188 <RUB>
A:Cross-references: EMBL:Z34465; NID:9600117; PIDN:CAA84230.1; PID:9600118

Query Match 10.6%; Score 784; DB 2; Length 1188;
Best Local Similarity 29.9%; Pred. No. 2,1e-25;
Matches 230; Conservative 59; Mismatches 359; Indels 122; Gaps 33;

```
QY 168 PTPKPPVVDDEAGSLDNGDFKVTTPDTSTQHNKVTSPKITTAKPIPRRESL-----P 243
DB 460 PTPHSPPAD-----DYVPTTPPVYVKGSPATSPSQVQPPAASPPPSLVKLSBPQ 510
QY 244 SDTSKETSILVNNKETTETKETTNNKOTSDGKEKTTSAKETOIEKTSKDLAPTSKY 303
DB 511 APVG-----SPPPVKTTSPAPIG-----SPSPPPVS 541
QY 304 -----LAKPTKAEITTKGPAALTTPKEPTTPPKEPASTTPKEPTTPPKISAPTTPKE 358
DB 542 SPPPVVSPPPPAVGVSPPEKSPPPPAVAVSPPPVKSP--PPPIVAVSPPPVKSP 599
QY 359 PTTKSAPTTP--KEPAPTTPKEPAPTTPKEPAPTTPKSAPTTTP--KEPAPT 415
DB 600 PPAVAVSPPPVKSPPPPTPVASPPPAVAVSPPPKSPPPPPTPVSSPPPEKSPPP 659
QY 416 PKKPAPTTPKEPAPT-----PKE-----PTTPPKEPAPTTPKEPAPTTPKEPAPTAP 466
DB 660 PAKSTPPEEYPTPTSVKSSPPEKSLPPTILPSPPEKPTPSTPKP--PSSPEK 718
QY 467 APITPKEPAPTTPKEPAPTTPKEPAPTTPKSAPTTTPKSAPTTTPKSAPTTTPK 526
DB 719 SP--PKPEVSSPQT-----KSSPPPAVSSPPTPVSSPPALAVVSSPVKSSP----- 768
QY 527 EPSPTTPKEPAPTTPKEPAPTTPKPAPTTPKPAPTTPKEPAPTTPKPAKPAKPA 586
DB 769 PPAVLSPPPAPOVKSSPPVQVSSP--PPAPKSSPLAPVSSPPVQEKSSPPAPLSSPP 827
QY 587 TTPKETAP-----TTP-----KKLTPTPEKLAPTTPKEPAP-----TTPPELA 633
DB 828 LAPSSPPAHVAVSSPPVAVSSPPPAVSSPPLPKPASPAPAHVSSPPEVAVSP--PA 885
QY 634 PTPPEEPAPTTPKAAAPNPKE--PAPTTPKEPAPTTPKPAPTTPKEPAPTTPKGAAPT 692
DB 886 PTVVISP--PSEPKSSPPPTVLSLPPPIVAVSSPPPAVSSP--PWTPKSSPPPVVSSPPT 943
QY 693 LKE-----PA-----PAPKELAPTTTKE-----PTSTSDKPAPTTPKGAAPTTP 738
DB 944 VKSSPPPAVAVSSPPATPKSSPPPAVNLPPPEKSSPPPTPVSSPPPA--PKSSPPAP 1000
```

A:Accession: T30826
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2187 <TOT>
A:Cross-references: EMBL:048363; NID:g1666688; PID:g1666689; PIDN:AAH18732.1
C:Genetics:
A:Gene: Nacc
A:Map position: 10
A:introns: 24/1; 1996/1; 2024/3; 2050/3; 2099/3; 2142/3; 2183/3
A:Note: differential splicing converts alphaNAC into a tissue-specific DNA-binding activity
C:Keywords: alternative splicing; DNA binding; transcription factor

Query Match 11.0%; Score 812; DB 2; Length 2187;
Best Local Similarity 27.4%; Pred. No. 2.7e-26;
Matches 311; Conservative 128; Mismatches 449; Indels 246; Gaps 51;

QY 85 NPTSPSSKKAPPPGASOTIKSTKRSPKPKKTKKVISEETEBHEHSENOESS 144
DB 746 SPT-PPSSKGAIVPSTGA-----PPSPKGAIVPRESSISSKQVPAEILPFO 792
QY 145 SSSSSSSSSITIKIKSSKSAANRELÖKKLVKDNKKKTKKPPVVDAGSGLDN 204
DB 793 KPEVYATSLAISOVPKVDPLMSD-----VTPSPKTKSATAVPK----- 833
QY 205 GDFKVTTPDTSTTOHKK-----VSTSPKITTKAPINRPSLPNDSOTKETSL--TVNKE 257
DB 834 -----DTSATLSLKVPAVATSLSPKAPVAPSNEATVPEITSLKNAIAAATPK 885
QY 258 T-----TYETKETTNNQSTSDGKEKTSKENSQISKTSKDLAPSKVLAKPTPKA 311
DB 886 TLAATSPKVTSPSPQKTPSVSLKGPAMTSKAT---ELAASKDVSPQ--FPKRVPL 940
QY 312 ETATGKPAITTPKEPTTPKPEASTTPKEPTTITKSAPTTPKE--PAPTTTKSAPTTPK 370
DB 941 QHV---PPSPSPKSPVSDTLISGLTSPKGPATILAEPTVKKKSPKPAASKTPATPS 997
QY 371 -----EPAPTTKEPAPTTKPEAPTTKEPAPTT--TKSAT----- 406
DB 998 PEGVTAIVLEIPPCSKKAPTAAPKESASVSSSKRAPKTAIVSKETISGVTAVPLEISLP 1057
QY 407 -----TPKEPAPTTPK--PAPTTKPEPAPTTKPEPTTPKPEPAPTTKEPAPTTPK 457
DB 1058 LKETSATGEGKSASSPKSPKATGPK---TPPGVTAIVPEILSPKETEQNMTPTNE 1114
QY 458 PAPTAAPKAPATT--KPEAP--TTPKPEPAPTTKESPTTPK-----EPAPTTT 506
DB 1115 SLAASSQKRSBPSTVSKETPPGGVTAIVLEIPSAPOKAPTAIVPKOIPPEDAVTTIAGS 1174
QY 507 PTTTKEPAPTTT--KSAPTTPK-----EPSPTTTKEPAPTTKPEPAPTTPKKAPATT 556
DB 1175 PLSPKKASKTAAPKAPKAPTAIVSGEISPSPKTSTAAPKENSATILPKRSPKTA 1234
QY 557 -PK-----PAPTTKPEAP-----TTTKKAPAPTAAPKPEPAPTTKETAAPT 596
DB 1235 APKETPATSSGVAIVSEISPSPTPASKGVPTLTPKGAIVNLAIE--SPASPKVAKTA 1293
QY 597 PKKLTPTPEKLAIVTPEK--PAPTTPELAIVTPEEPTPT-----PEEAPPTTPKA 647
DB 1294 APEESTTP-----SPOKIPIKVAIPKEASATIPSKTTPKTAIVKETSASSEGVTAIVLE 1347
QY 648 AAPNTPEPAPTTTPEK--PAPTTKPEPAPTTKETAAPTTPKGTATVTTKEPAPTT----- 701
DB 1348 IPPSBRKAPKTAAPKEPPAPPS--DEGATTAIVQIPPSRKSCKAAGSKE--TPTTPSEGV 1404
QY 702 -----KKPAKELAPTTTKEPTT-----TTSKPAPTTPKGTAPTTTPEK 740
DB 1405 TAAPLEITISSKTSKMASPKETILVTPSSKTLQTVGKETSLBGATAVPLEIPPSHKA 1464
QY 741 PAPTPPKK--PAPTPPKGTAPTTTKEPAPTTTTPKAPKLAATTTTGKPTSTTSKDPAPTT 799
DB 1465 PKTVDPKGVPLTPSPK--DAPTTLAE--SPSSPKK--APKTAAPPSER--VTTVPPEKPA--TPQ 1519

QY 800 KETAPTTKPEPAPTTTPKPPAPTTPEMP-----PPTSEVSTPTTK-----EPTTIHKS 848
DB 1520 KASGTTASKVAVPAEPTQEVAVSSKETPTVTPAVPVVKNPSSHKRTSTILKEKAPATLPPS 1579
QY 849 PDESTPELASEPTPKALENSPKR--PGVPTTKTPAATKPEMTTAKOKTTERDRLRTPEPT 907
DB 1580 PTKSPKIPSSKKAPRT--SAPKEPPASPIK--PVTT--SLAQTAAPSLQKAPSTTIIPKEN 1634
QY 908 TAAPKMTKETATTTKTESKITATTTQVSTTTQDTPPEKITTKTTLAPAVTTTKK 967
DB 1635 LAAPAV-----LPVSSKSPAPPARASASLSPATAPOTAPEKATTTISCKRAAATETPIER 1690
QY 968 ITTTEIMNKPEETAKPKDRATNSKATTPKPOKPTKAPKPTSTKPKTPMVRVKPKTPT 1027
DB 1691 STASLEAPKETSE-----TSVKVLMSSP-----PKKASSKRASTLP----- 1730
QY 1028 PRKMTSTPELNPTRIAEAMLOTTTPRNPOTPNKLVVNPKSPDAGAGAGETP 1081
DB 1731 -----ATLTPSIKEASVLS-----PTATSSGKDSHISFVS--DACSTGTTTP 1770

RESULT 6
548478
glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast (Saccharomyces cerevisiae)
N:Alternate names: extracellular glucosylase; mucin-like protein MUC1; protein YIR01
C:Species: Saccharomyces cerevisiae
C:Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text-change 12-Nov-1999
C:Accession: 548478; A26877; B26877; S27281; J06123
R:Rowley, K.
Submitted to the EMBL Data Library, October 1994
A:Reference number: 548478
A:Accession: 548478
A:Molecule type: DNA
A:Residues: 1-1367 <ROW>
A:Cross-references: GB:247047; EMBL:238061; NID:g603997; PID:g763364; GSDB:GN00009;
R:Yamashita, I.; Nakamura, M.; Fukui, S.
J. Bacteriol. 169, 2142-2149, 1987
A:Title: Gene fusion is a possible mechanism underlying the evolution of STAL.
A:Reference number: A91831; MUID:87194600
A:Accession: A26877
A:Molecule type: DNA
A:Residues: 1-242 <YAN>
A:Cross-references: EMBL:M16164; NID:g172522; PIDN:AAA5014.1; PID:g172525
A:Accession: B26877
A:Molecule type: DNA
A:Residues: 762-1331 <YAO>
A:Cross-references: EMBL:M16165; NID:g172523; PIDN:AAA35015.1; PID:g172526
R:Pardo, J.M.; Ianez, E.; Zalacain, M.; Clares, M.G.; Jimenez, A.
FEBS Lett. 239, 179-184, 1988
A:Title: Similar short elements in the 5' regions of the STA2 and SGA genes from Sacc
A:Reference number: S27281; MUID:89031230
A:Accession: S27281
A:Molecule type: DNA
A:Residues: 1-31 <PAR>
A:Cross-references: EMBL:X13857; NID:g4551; PIDN:CAA32069.1; PID:g4552
R:Lambrechts, W.G.; Bauer, F.F.; Marmur, J.; Pretorius, I.S.
Proc. Natl. Acad. Sci. U.S.A. 93, 8419-8424, 1996
A:Title: Muc1, a mucin-like protein that is regulated by Mss10, is critical for pseud
A:Reference number: J06123; MUID:96323237
A:Accession: J06123
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-1367 <LAN>
A:Cross-references: GB:030626; NID:g1304386; PIDN:AAC49609.1; PID:g1304387
C:Genetics:
A:Gene: SGD:MUC1; STA2; MAL5; DEX2; SGD:S0001458
A:Cross-references: MITS:YIR019C; SGD:S0001458
A:Map position: 9K
C:Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosida
C:Keywords: glycosidase; hydrolase; polysaccharide degradation; transmembrane protein
F:5-21/Domain: transmembrane #status predicted <TM1>
F:1350-1366/Domain: transmembrane #status predicted <TM2>

Db 988 ETTAAPTETMYAPTEETTYAPAEETPYEPTETTYAPTEETTYAPTEETTY 1047
 QY 761 -----TLKEPAPTPPKP-----APKE---LAPT---TKGPTSTSDKAP 796
 Db 1048 ASTEETTYAPTEETTYAPAEETPYEPTETTYAPTEETTYAPTEETTYAPTE 1107
 QY 797 TTPKETAAPTPKEBAPTPPKAPPTPEPTTSEVSTPTTKKEPT--TIHKSPTDE 854
 Db 1108 ETTAAPAETPYEPTETTYAPTEETTYAPTEETMYAPTEETTYGPIETTY-APTAT 1166
 QY 855 ELSAEPKPA-LENSKPEG-----VPTTKPAATKPEMTTAKDKTERDLRTTP- 904
 Db 1167 YAPTEETPYAPTEETTYEPGETTYAPTEETTYAPTEETTYAPTEETTYAPTE 1226
 QY 905 ETTAAPKMTKETATTEKTESKITATTQVSTTTODTT-----PEKITP--- 951
 Db 1227 EETTYAP-----TEETTYEPTETTYAPTEETTYAPTEETTYAPTEETMYAP 1281
 QY 952 LKTTTLAPKVTYTTKTTTTEIMNKPPEYAKPKDRAT--NSKATTPKP-OKPTKAPK 1008
 Db 1282 TEETTYAPTEATYAPTEETTYEPTETTYEPTETTYAPTEETTYAPTEETTYAP 1341
 QY 1009 -----STKRP-KTMPRVKRPKTTPTPKMTSTWPELNPTRISAMIQTTTP 1055
 Db 1342 PYEPAESTVSTKPCNTBEETDEPTDEPTDE--PSEPTDEPTDEPTDEPTDE 1399
 QY 1056 --NOTPNSKLEVNPKSEDAG 1074
 Db 1400 CDNOGINGIGVENKVRNNAG 1420

RESULT 4
 T16251
 hypothetical protein F35A5.1 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 18-Feb-2000
 C:Accession: T16251
 R:Leinbach, D.
 submitted to the EMBL Data Library, January 1996
 A:Description: The sequence of C. elegans cosmid F35A5.
 A:Reference number: Z18485
 A:Accession: T16251
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1274 <LEI>
 A:Cross-references: EMBL:U46675; NID:g1166613; PID:g1166621; PIDN:AAB52641.1; GSPDB:GN00
 A:Experimental source: strain Bristol N2; clone F35A5
 C:Genetics:
 A:Gene: CSP:F35A5.1
 A:Map position: X
 A:Introns: 1272/2

Query Match 11.4%; Score 846; DB 2; Length 1274;
 Best local similarity 27.6%; Pred No. 6.5e-28;
 Matches 313; Conservative 119; Mismatches 455; Indels 248; Gaps 60;

QY 90 PSKKAAPPSGASOTIKSTTKRSP---KRP---NKKTKKYLESEITEHSVSENQES 142
 Db 185 PSFKKAAPSEKNDPIVPPIPKNPAKKMKRPWEDDEVPEETKEPPATRKYPALKKKRP 244
 QY 143 SSSSSSSSSSTIMIKSSKNSAANRELOKKILKYKDNK-KNRTKK-KP-----TP----- 190
 Db 245 STSVKAVSDPSPTKKY-----PVKKEPEVPETPIKNPKKKRPWEDETPVEEVK 294
 QY 191 KPPVNDAGSGLDNGD---FKVTPPDSTTOHNKYSTSPKITYTAKPI--NPPRSLPPNSD 245
 Db 295 EPPVPEKKAAPVLKKKDPAPAAKADSPSKAPKAPKVEPSPVPPVKNPVPKPKKPPME 354
 QY 246 TSKETSILYANKETVETKETTNTKOTSDGKEKTTSAKETOSIEKTSKADLAFTSKYLA 305
 Db 355 VDDEPAEEVKKDSAPKPKPVLRKKEPEPSSSTPPSSDPSPKKAAPAVKVRDSDSPKKAPPL 414

QY 306 KPIPKAE-----TTTKGA-----LTPKEPPTTP-----KEPASTTPK 340
 Db 415 QADPKAQEVPPTPVKNKPVKKRPWEDDEVPEEVQAPAPAKPTPVLRKKEPAKKDPA 474
 QY 341 EP-----TP--TTISAPPTPKPAPTTTKSAPTTPKPAPTTTKEPA----- 381
 Db 475 KPATSKPTPEPKKDPVKPDDSSPKVAAKPSDAQAA--TPVKNPVKKRPWEDDETPA 533
 QY 382 -----PTPKPAPTTTKEPAP-----TTKSAPTTKEP-----APTTPKPAPTTP 424
 Db 534 DVSKPPTDAKKTSLAKDPAKESILKPAKDPAKAPKPDSPKKVAPTPAPKKTVPVLA 593
 QY 425 KE-----PAPTPKEPPTTPTTKKEPAPTTKKEAPTPPK--EPAPTPAKKPA--TTPK 473
 Db 594 KKEPAGADSKTKEPEKSKRDPSPKAVPAKVPPTTEVAAPVKKKEPISKRDPAPK 653
 QY 474 PAPTPKEPAPTTTKEP-----SPTPKPAPTTTKSAPTTPKPAPTTTKSAPTTP--KEP 528
 Db 654 AEPNSPVVP--PTVKNPVKKMKRPWEDDDAPAKPVSLPEPEK--TVLAKKAPTKDSEA 711
 QY 529 SPTTKEPAPTPPK--EPAPTPKKPA--TPPKPAPTTPKPAPTTTKKEAPTPA--PRE 583
 Db 712 AADPVSGPSKDKLAKKAPVKRDPSPMKAVPIKPAKT--EVPPAVVKKPEPAKSD 769
 QY 584 PAPTPKEAPTPPKKLTPTTP-----EKLAPTPKPAPTTPPEELAPTTPEEP-----T 633
 Db 770 PSKKAAP--AEPNSP--VPEPTPVKNPVKKRPWEDDDAPAEVNVPEPEKKTVPYLA 826
 QY 634 PTPKEPAPTPKKAAPNTPKKEPAPTPPKEPAPTT--PKEPAPTPKETAAPTPPKGAPTT 692
 Db 827 PVKPRDPSPKKAVPAKPSKTDAPVSVKKPEVSKRKEPSKKAEPNSPVVP-----PTP 882
 QY 693 LKPAATTPKPKPAKPLAT--TKKEPTSTSDKPAPTTPKGAAPTPPKPAPTTPKPA 751
 Db 883 VKNPVKKW--KPWEDDEDETEEVKKPSE--PEKKTVPVLA--KEPEKPD--APKVAAPKPD 937
 QY 752 TTPKGAPTTLKEPAPT-----TPKKRAP-----KELAPTTTGKPS-----TTSK 793
 Db 938 PSFKKAVPE--KEPAKVAAPKPDLSKKAIPANTQAPPTPVKNPVKKMKRPWEDDE 995
 QY 794 P-----APTPKET-----ADTPKEPAPTPPKRAPPTPEPTTSEVSTPTTKEP 842
 Db 996 PAEPVSAPEPEKKTVPVLAAPKAPKPD--SPKKAAPVAAK--PDKPIEV--PPTPVKNP 1050
 QY 843 TTIHKS-----DESTPELSA--BP-----TRKALNSKEPEV-----PTTKPA 881
 Db 1051 VKKMKRPWEDDEPSPVSAPEPEKKTVPVLAAPKATPKATKPSDAADPVSGPTSKDPK 1110
 QY 882 AT-----KPEMTTAKDKTERDLRTTPET--TAAPKMTKETATTEKTESKITATT 933
 Db 1111 LSKKAPVEKRPPTTDPKDKLKPSPAKRDEKAPDEPAARKKKRPVWDDDEDEAFTVPA 1170
 QY 934 TOVSTTODTTPEFKITTLKTTTLAPKYVTTTKKTTTTEIMNKPPEYAKPKDRATNSKAT 993
 Db 1171 PSKKPDTEPADPLG-----GPKTKDPK-----LNKKAAPAEKPEK----- 1206
 QY 994 TPKPOKPTAPKPKSTKPKMTMPVRKPK-----TTTPPKMTSTMP 1036
 Db 1207 -PKPEKVSKEPKPIEPPRP--AAPKMKRPWEDDDPEADFTMPAPKPKPTEDP 1259

RESULT 5
 T30826
 nascent polypeptide-associated complex alpha chain, muscle splice form gp220 - mouse
 N:Alternate names: alpha-NAC protein
 C:Species: Mus musculus (house mouse)
 C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 24-Nov-1999
 C:Accession: T30826
 R:Yotcov, W.V.; St-Arnaud, R.
 Genes Dev. 10, 1763-1772, 1996
 A:Title: Differential splicing in of a proline-rich exon converts alphaNAC into a mus
 A:Reference number: Z20869; MUID:96312450

Query Match 12.8%; Score 950; DB 2; Length 1664;
Best Local Similarity 31.0%; Pred. No. 4.5e-32;
Matches 312; Conservative 104; Mismatches 361; Indels 230; Gaps 55;

```
QY 303 VLAKPPP-KAEITTTGKAPLITTKKEPPTTPKKEPAPTTPKKEPPTTKKSAPTTPKKEPAPT 361
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 758 VVIOPAPIKAAADDEPIPTDPSDEPPS-----DEPPS--DEPTPSDEPTPSD 804
QY 362 TKSAPTTPKKEPAPTTPKKEPAPTTPKKEPA-----PTTKKEPAPTTPKSAPTTPKKEPAPTTPK 417
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 805 EPIPTSPTEPIPTDTPSPDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 864
QY 418 KPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEP 476
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 865 SDEPTPSDEPTPS--DEPTPS--DEPTPS--SDEPTPSDEPTPSDEPTPSDEPTPSDEPTPS 919
QY 477 TTPKKEPA--PTTKKEPSPT-TPKKEPAPTTPKS--APTTPKKEPAPT--TTKSAPTTPKKEPS 529
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 920 SDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 979
QY 530 ----PTTKKEPAPT-TPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEP 584
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 980 PSDEPTPSDEPTPSSEPTPEEPIPTDPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDE 1039
QY 585 APT--TTPKETAPTTPKKLPTTPKELAPITTPKKEPAPT--TPEELAPITPEEPTPT-TP 637
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1040 TPDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDE 1094
QY 638 EEPAPTTPKKAAPTTPKKEPAPTTPKKEPA--PTTKKEPAPT-TPKETAAPTTPKKTAPTTPK 694
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1095 EEPITPTDPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSD 1154
QY 695 EPAPTTPKKEPAPKELAPITTPKKEPT-STTSOKPAPTTPKGTAPTTPKKEPAPTTPKKEPAPT 753
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1155 EPTPS--DEPTPSD-EPTPSDEPTPSSEPTPEEPIPTDPSDEPTPSDEPTPS--DEPTPS- 1208
QY 754 PKGTAPTTPKKEPAPT-TPKKAPKELAPITTKGPTSTSKPAPT--TPKETAAPTTPKE 809
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1209 ----DEPTPSDEPTPSSEPTPEEPI-----PTDPSDEPTPSDEPTPSD-EPTPSDE 1253
QY 810 PAPTTPKKEPAPT-TPETPTTSEVSTPTTKKEPTTIHKSPDSSTPELSAAPTTPKALENS 868
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1254 PTPS--DEPTPSSEPTPEEPIPTDPSDEPTPSDEPT--PSDEPTP--SDEPTPSDEPT 1305
QY 869 PKKEGPTTPKPAAT--KPKMTTAKDKTERDLKITPETTTAAPKAKTKETATTEKITE 926
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1306 SDEPTPSDEPTPSSEPTPEEPIPTDPSDEPTPSD--EPTPSDEPTPSDEPTPSDEPTPS 1361
QY 927 SKIATITTOVSTTODTTPFKITTLKTTTLAPKVTTTKKITTTTEIMNKPDEETAKPKDR 986
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1362 DEPTPSDEPTPSSEPTPEEPT--TTPTPTPTTT-----PTSG 1396
QY 987 ATNSKAT-----TPKPOKPTKAP--KKPTSTKKPKTMPVRKKPTTPPKK----- 1031
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1397 SCGCGGGGGGGGGGCTVPSPTPTPSKPTAP--TEIEEPTSDVGAIGCEHRA 1453
QY 1032 -----TSMPELNPSTRIAAEALQTTTRNCPNSKIIVNPKSDAGABEETPHM- 1083
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1454 YLRGYPDGSFPERNITRAEAAVIF-----AKLL-----GADSSGAASAPISD 1498
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1084 LLRHP-----VFMP-----VTPDMDYLPVYNOGI----- 1109
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1499 LAOTHHMAAIAIKFATSGLEFGYDGTFFKPOQNTITRAEFAIVLHFLTKVAGQELMSKLA 1558
QY 1110 ---IINPALSDETNICNS---KPYDGLTTL-----RNGTLVAPRGHYFWMLSPSPSPS 1156
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1559 TIDISNPFDD---CVGHAQEFLEKLTSLGISGYEDGT-----FKPON 1600
QY 1157 PARITEWGIPIPDIVFTRCNCEGKTFPKD--SOVWRTNIDIKD 1201
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1601 YIKRSESV---ALINRALERGLNGAKLPDPVNESYMAF-GDITMD 1642
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 3
T31108
cyst germination specific acidic repeat protein precursor - phytophthora infestans
C:Species: Phytophthora infestans (potato late blight agent)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C:Accession: T31108
R:Goenhardt, B. EMBL Data Library, April 1998
Submitted to the number: 220986
A:Reference number: 220986
A:Accession: T31108
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1489 <GOE>
A:Cross-references: EMBL:AF061185; NID:g3851513; PID:g3851514; PIDN:AAC72308.1
C:Genetics:
A:Gene: car90

Query Match 11.7%; Score 863.5; DB 2; Length 1489;
Best Local Similarity 31.7%; Pred. No. 1.4e-28;
Matches 368; Conservative 55; Mismatches 509; Indels 229; Gaps 51;

```
QY 121 TKKVISEETTEHVSSENQSSSSSSSSSTTWIKIKSKNSAANRELQKKLVKDK 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 282 TPTVGYSTEEETEGQHVYGVYPSDETEAPTEGTYV--VPREETAPPS-----DTYAP 334
QY 181 KNTTKKKKPKPKPVYVDEAGSGLDNGDFKVTTPDTST-----QHNKYSTPKITAKPI 234
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 335 REVTPYAPTEKPYDEETTYVVEETTYAPTKSTNAPTERMHYAHLEKPCDIEVIMYAPT 394
QY 235 NRPASLP-----PNSDT-----SKETSLVANKETVETKETT---TNKOT 272
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 395 EETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTE 454
QY 273 STDGKEKTSIAKTOGSIKTSADLAPTSKVLAKPPTPKMETTKGALTPKKEPPTTPK 332
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 455 YAPTEETTYAPTEETTYAPTEETTYAPTKETTYAPT--EETTYASTEETTYAPTEETTY 511
QY 333 EPASTTPKEPTPTTIKSAPTTPKKEPAPT--TTKSAPTTPKKEPAPT--TTPKKEPAPTTPKKEPA 389
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 512 APAEETPYEETET--TYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAEETTYEPT 570
QY 390 PTTTKEPAPTTPKSAPT-----TPKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPT 439
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 571 EETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTE 629
QY 440 PKEP-----AP-----TTKEPAPTTPKEP-----APTAPKRAPPTTPKKEPAPT 476
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 630 TYAPTEETTYASTETTYAPTEETTYAPAEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAP 689
QY 477 TTPKKEPAPT--TTPKESPTTPKEP-----APT--TTKSAPT-----TTKEP----- 513
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 690 TEETTYAPTEETTYAPAEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTE 749
QY 514 --APT--TTKSAPT--TPKESPTTPKEP-----APTTPKEPAPTTPKKEPAPTTPKKEPAPT 562
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 750 TYAPTEATTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAP 809
QY 563 T--TPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPT 620
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 810 TEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTE 867
QY 621 PEEAPAPT--TPEEPTTPTEEP-----APT-----TPKAAAPTTPKKEPAPTTPK 662
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 868 EETTYAPTEETTYAPTEETTYAPTKETTYAPTEETTYASTETTYAPAEETTYAPAEETTYAP 927
QY 663 EP-----APTTPKEPAPTTPKETAAPTTPKGTAPT--TTKEPAPTTPKKEPAPT--ELAPT- 712
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 928 EPTTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAEETTYEPEETTYAPTE 987
QY 713 -TTKEPSTTSIDKRAPPTPKGTAPTTPKKEPAPTTPKKEP-----APTTPKGTAPT----- 760
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

J. Clin. Invest. 87, 77-82, 1991
 A:Title: Human bronchus and intestine express the same mucin gene.
 A:Reference number: A61257; MUID:91086481
 A:Accession: A61257
 A:Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 191,192-1948,1952-1954 <JAN>
 A:Experimental source: bronchus
 R.Xu, G.; Huan, L.; Khatri, I.; Sajjan, U.S.; McCool, D.; Wang, D.; Jones, C.; Forstner, A.; Blochem. Biophys. Res. Commun. 183 821-828, 1992
 A:Title: Human intestinal mucin-18 protein (MLP) is homologous with rat MLP in the C-terminus
 A:Reference number: P00328; MUID:92198477
 A:Accession: P00328
 A:Molecule type: mRNA
 A:Residues: 2328-2468 <XUG>
 A:Cross-references: GB:M6523
 A:Experimental source: small intestine
 A:Accession: P00329
 A:Molecule type: protein
 A:Residues: 2328-2342, K, 2344-2354 <XUG1>
 A:Gene: GDB:MUC2
 A:Cross-references: GDB:120203; OMIM:158370
 A:Map position: 11p15.5-11p15.5
 C:Superfamily: von Willebrand factor; von Willebrand factor type A repeat homology; von F;2766-2834/Domain: von Willebrand factor type C repeat homology <VWC>

Query Match 14.7% Score 1090; DB 2; Length 3020;
 Best Local Similarity 27.4%; Pred. No. 1.4e-37;
 Matches 378; Conservative 107; Mismatches 514; Indels 382; Gaps 51;

QY 19 TCNCDY--NCOHWECPCDPKRVCTAELSCGRCFESFERGECDDACCKKYDKCCPDY 76
 Db 1079 SCSDIDTGDGDECECSAASVSAQECTKEGAC-----VFMTPLD-CIFEDYIN---PRH 1128
 QY 77 ESFCAEVHNPTSPSSKAPAPPSGASQTI-----KSTTKSPK--PNKKTATKVIKVESE 129
 Db 1129 E--CEMHYECGNSEFTCTINGIHNSISVTLGCGYPCPKDRPIYEDLKCCTATADK 1186
 QY 130 -----ITEEH-----SVSNOSSSSSSSSSSSTIWKIKSSKNSAANRELOKRLKVDNKK 181
 Db 1187 CGCYVEDTHHPGASVTEETCKSCVCTNSGVVCPREBK-----IL 1229
 QY 182 NRTKKRPPTKRPVVDAGSLDNGDF--KVTPPDISTQH--NKVSTSPKTTAKPIN 235
 Db 1230 NQOQ-----DGAFCYWEICGPNGTVEKHFNICSIITRPSLTITFTTI 1271
 QY 236 PRSLPNSDTSKETSILYVKETVEKET--TTNK-----QTSDDGKE 278
 Db 1272 TLPTTP-----TSFTTTTTTTTTTSSVLTSTPKLCLCNSDWINEHPSGSDGDR 1323
 QY 279 KITS--AKETOSIEKTSAD----- 296
 Db 1324 EPPDVGAGAEDEICRSVADPHLSLBOHOKVOCDSVGFICKNEDQFNGPFGLCYDK 1383
 QY 297 -----LAPTSKYLAKPTPKAETTTKGPALTTPKEPTTPPKAPASTTKEPTP 344
 Db 1384 IRVNCMPMDKCTITPSPTTPSPPTTTTTLPTTSSPTTTPPTTTPPTTTPPTTSPPT 1443
 QY 345 TTKKSP--TPKEPAPTTTKASATPKPEPAPTTKEPAPTTKEPAPTTTKAPATTKS 403
 Db 1444 TTTTPPTTTPSPSPISSTTTTTPPTTSP--PTTSSPTTTPSSPT--TTTTTTPPTTTPS 1501
 QY 404 AP--TPKEPAPTTPKKAPATTTKEPAPTTPKPEPTTPPTTKEPAPTTKEPAPATTA 462
 Db 1502 PPTTTPITTPASTTTLPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTT 1560
 QY 463 PKKAPATTPKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTT 521
 Db 1561 TSPSPPTTTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTSP 1620

QY 522 P--TPKEPSPTTTKEPAPTTPKPEPAPTTPKKAPATTPKPEPAPTTTKEPAPTTKAPPA 580
 Db 1621 PTTTPITTPPTSTTLPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPT 1680
 QY 581 PKKAPATTPKEAPATTPKKLPTTPEKLAATPEKAPATTPPEELAPTTPEEPPTPEEP 640
 Db 1681 PSSPTTTP--SPTT-----TMTTPS--PTTSSPTTTPPTTTPSSSTTP--SPTTTPMTTP 1731
 QY 641 APPTPKAABNTPEKAPATTPKPEAPPTPKPEAPPTT-----KETAPTP----- 685
 Db 1732 SPTTTPPTTTPMTTLPTTTPSSPLATTLPLPSITTPPSFSTTPPTTPCVLCNMWTGM 1791
 QY 686 --KG----- 687
 Db 1792 LDSGKRNFKRGDTELICDVCSPGMANISCRATWYPDVPIGOLQIYVCDVSVGLICK 1851
 QY 688 -----IAPTLKAPATTPPKKAPKELAPTT--TKRP 717
 Db 1852 NEDQKRGCVIAPAFCLNVEINVOCCCVQPTTM--TTTTTENPPTPPTITTTTTPPTP 1908
 QY 718 TSTTSDKP---APT--TPKGTAPTPKEPAPTTPKPEPAPTTPKGTAPTLKPEAPTPPK 772
 Db 1909 TPSTQSPNGLQAPPTPISTTTTTPPTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTT 1967
 QY 773 PAKKELAPTTTGPST--TSDKAPATTPKEIAPTTPKPEAPTPPKP-----APTPPE 824
 Db 1968 PTTVLTITTTTTPPTTTPPTTTPSTKSTVTPTTTTTPVATPPTPPTGQTPMIPISTTTTTP 2027
 QY 825 TPEPTTSEVSTPTTTPK-----EPTTIHKSPEDESTPELSAEPKALENSAPKRPPTTKT 879
 Db 2028 TPPTTGSPTPTPTHTSTAPIALTISNPPESSTQTSISSTPLTESTTTLSTLPPAIE 2087
 QY 880 PAKTKEMTTAKDKTERDLRTPEETTPAABKMTKEATTEKTESKITATTOVST 939
 Db 2088 MNTAPSTPTAPPTTSGHTLSPPSTTSPGPTGTTT--GSSAPPTSTVQTITTT 2145
 QY 940 TQDPTPEKITT---LKTTLAPKVTYTKKITTITTEINKPBE----- 979
 Db 2146 SAMTPTPLPLSTPLSIRITGLAPRYSVALICVLDITYAAGEEYVNGTVDCYFNCS 2205
 QY 980 -----TAKPKDRAINSKATTPKPOKPTKAPKPTSTKPKTPMPVRKPKPTPT 1027
 Db 2206 ISCTLEFYNMSCPSYSPSTPPTPSK--STPTPSKPSSTPSKPTPGKRPPECPDPPR----- 2260
 QY 1028 PKKMTSTPELPTSRIDAMLOTTTRNONTNKSILVNP-----KSDAG 1074
 Db 2261 -----QENETWMLDCDFH--ATCKNTNVEIYKVECEPPMPPTCSNGILOPVRVEDPD 2310
 QY 1075 G-----AEGTPHML-----RPHVFMPEVTPDDMDYLPRVPOGII 1111
 Db 2311 GCMWHMBCDCYCTGMD--PHVTFEDGLYSYGNGCTYVLVEISPSVD-----NFGVI 2363
 QY 1112 N 1112
 Db 2364 D 2364

RESULT 2
 118262
 S-layer protein - Clostridium thermocellum
 C:Species: Clostridium thermocellum
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #ext_change 15-Oct-1999
 C:Accession: T18262
 R:Fujino, T.; Beguin, P.; Aubert, J.P.
 J. Bacteriol. 175, 1891-1899, 1993
 A:Title: Organization of a Clostridium thermocellum gene cluster encoding the cellulose
 e.
 A:Reference number: Z18847; MUID:93209931
 A:Accession: T18262
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1664 <FUU>
 A:Cross-references: EMBL:X67506; NID:g296879; PID:g296881; PIDN:CAA47841.1

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:18:10 ; Search time 82.85 Seconds
(without alignments)
1268.810 Million cell updates/sec

Title: US-09-556-246-1_COPY_25_1404
Perfect score: 7410

Sequence: 1 QDLSCAGRCGEGYRDATC.....ARATITRSGQTLKRWYNCP 1380

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: PIR_68:*
2: PIR1:*
3: PIR2:*
4: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being predicted, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1090	14.7	3020	2 A43932	mucin 2 precursor, S-layer protein -
2	950	12.8	1664	2 T18262	cyst germination s
3	863.5	11.7	1489	2 T31108	hypothetical prote
4	846	11.4	1274	2 T16251	nascent polypeptid
5	812	11.0	2187	2 T30826	glucan 1,4-alpha-g
6	809	10.9	1367	1 S48478	extensin-like prot
7	784	10.6	1188	2 S49915	high molecular mas
8	695	9.4	1151	2 T18535	hypothetical prote
9	676.5	9.1	1229	2 T23697	mucin 1 precursor,
10	672	9.1	1344	1 A35175	elastic titin - hu
11	666.5	9.0	7962	2 T38346	mucin MUC5B, trach
12	663	8.9	3570	2 T45025	hypothetical prote
13	658.5	8.9	3507	2 T34513	nucleolar phosphop
14	643.5	8.7	990	2 T51618	hypothetical prote
15	633	8.5	489	2 T11622	extensin class 1 p
16	632	8.5	761	2 C84672	hypothetical prote
17	629	8.5	6642	2 T29757	protein UNC-89 - C
18	627.5	8.5	971	2 T19431	hypothetical prote
19	622.5	8.4	839	2 F75518	hypothetical prote
20	607.5	8.2	801	2 T29018	gene LF3 protein -
21	607.5	8.2	924	2 S27293	larval glue protei
22	605	8.2	379	2 S50155	hypothetical prote
23	591.5	8.0	2232	2 T34434	ascites siatoglyco
24	568.5	7.7	1630	2 A53577	hydroxyproline-ric
25	559.5	7.6	350	2 S22456	polyphenolic adhes
26	559	7.5	875	2 S23760	hypothetical prote
27	556.5	7.5	856	2 T16543	hydroxyproline-ric
28	551	7.4	620	2 S06733	neurofilament trip
29	551	7.4	1087	1 QPM5H	

30	545	7.4	873	2 A47283	calphotin - fruit
31	542	7.3	369	2 S20500	hydroxyproline-ric
32	540	7.3	1459	2 T32271	hypothetical prote
33	538.5	7.3	416	2 U00465	extensin precursor
34	537	7.2	1072	1 A37221	neurofilament trip
35	532	7.2	756	2 T27642	hypothetical prote
36	530.5	7.2	813	2 S70795	vesa protein precu
37	530.5	7.2	1162	2 JH0557	calcium-binding pr
38	530	7.2	865	2 A47282	hydroxyproline-ric
39	522	7.0	328	2 J00985	neurofilament trip
40	521.5	7.0	854	2 S02003	membrane glycoprot
41	518	7.0	866	2 T45462	hypothetical prote
42	518	7.0	1611	2 T38236	tegment protein 2
43	515.5	6.9	3534	2 T42567	mucin-like glycopr
44	513	6.9	1832	2 T31113	membrane glycoprot
45	512.3	6.9	867	2 T45463	

ALIGNMENTS

RESULT 1
A43932
mucin 2 precursor, intestinal - human (fragments)
N:Alternate names: mucin SMUC-41
C:Species: Homo sapiens (man)
C:Date: 10-Mar-1993 #sequence_revision 12-Apr-1996 #text_change 05-Nov-1999
C:Accession: A49963; A45106; B45106; A43932; B33532; A61257; P00328; P00329
R:Gum Jr., J.R., Hicks, J.W., Toribara, N.W., Siddiki, B., Kim, Y.S.
J. Biol. Chem. 269, 2440-2446, 1994
A:Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of t
A:Reference number: A49963; MUID:94132002
A:Accession: A49963
A:Molecule type: mRNA
A:Residues: 1-639 <GU>
A:Cross-references: GB:L21998
R:Gum Jr., J.R., Hicks, J.W., Toribara, N.W., Rothe, E.M., Lagace, R.E., Kim, Y.S.
J. Biol. Chem. 267, 21375-21383, 1992
A:Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both up
A:Reference number: A45106; MUID:93016075
A:Accession: A45106
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 626-1895 <GU>
A:Cross-references: GB:M94131; NID:g186395; PIDN:AAA59163.1; PTD:g186396
A:Note: sequence extracted from NCBI backbone (NCBIP:116706)
A:Accession: B45106
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 2037-3020 <GU>
A:Cross-references: GB:M94132; NID:g186397; PIDN:AAA59164.1; PTD:g186398
A:Experimental source: colon
A:Note: sequence extracted from NCBI backbone (NCBIP:116698)
R:Toribara, N.W., Gum Jr., J.R., Guhan, P.J., Lagace, R.E., Hicks, J.W., Petersen,
J. Clin. Invest. 88, 1005-1013, 1991
A:Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polym
A:Reference number: A43932; MUID:9158717
A:Accession: A43932
A:Molecule type: DNA
A:Residues: 1343-1350, 'L', 1352-1411, 'S', 1413-1448, 'P', 1450-1503, 'T', 1505-1915 <TOR>
A:Cross-references: GB:M74027; NID:g188863; PIDN:AAA59875.1; PTD:g188864
A:Note: sequence inconsistent with the nucleotide translation
A:Note: sequence extracted from NCBI backbone (NCBIN:55749; NCBIP:55750)
R:Gum, J.R., Byrd, J.C., Hicks, J.W., Toribara, N.W., Lamport, D.T.A., Kim, Y.S.
J. Biol. Chem. 264, 6480-6487, 1989
A:Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and evi
A:Reference number: A33532; MUID:89197956
A:Accession: B33532
A:Molecule type: mRNA
A:Residues: 1316-2193 <GU>
A:Cross-references: GB:M22405; NID:g188873; PIDN:AAA36334.1; PTD:g188874
A:Experimental source: Intestine
R:Jany, B.H., Gallup, M.W., Yan, P.S., Gum, J.R., Kim, Y.S., Basbaum, C.B.

DB 821 ADDXRPXPG-----STAPXAHGVT-----SAPDXRPXGSGTAPXAHGVTSAVDXRPX- 868
 QY 824 KTKKAPKKTSTKPKPTMPRVKRPKPTTPRKMTSTMPELNP-SRIAEAMLOTTTRPNQ 882
 DB 869 -GSTAXAHGVTSAVDXRPX-XPGSTAPXAHGVTSA-PDXRPXGSGTAPXAHGVTSAVDX 924
 QY 883 TP 884
 DB 925 RP 926

RESULT 15
 US-08-276-967-2
 ; Sequence 2, Application US/08276967
 ; Patent No. 5851817
 ; GENERAL INFORMATION:
 ; APPLICANT: Hardy, Daniel M.
 ; APPLICANT: Galters, David L.
 ; TITLE OF INVENTION: Species-Specific Egg-Binding Proteins of
 ; TITLE OF INVENTION: Sperm
 ; NUMBER OF SEQUENCES: 8
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Arnold, White & Durkee
 ; STREET: P. O. Box 4433
 ; CITY: Houston
 ; STATE: Texas
 ; COUNTRY: USA
 ; ZIP: 77210-4433
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/276,967
 ; FILING DATE: Submitted Herewith
 ; CLASSIFICATION: 530
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Kitchell, Barbara S.
 ; REGISTRATION NUMBER: 33,928
 ; REFERENCE/DOCKET NUMBER: UTSD:418\KIT
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 713-787-1400
 ; TELEFAX: 713-789-2679
 ; TELEX: 79-0924
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2476 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-276-967-2

Query Match 8.4%; Score 419.5; DB 2; Length 2476;
 Best Local Similarity 32.5%; Pred. No. 8.5e-19;
 Matches 163; Conservative 43; Mismatches 177; Indels 119; Gaps 31;

QY 406 PREPAPPTPKETAPPTPKKLPTTPPEKLAPTTPPEKAPPTPEELAPPTPEEPPTTPPEEP 465
 DB 313 PSEISVSTSEKVPAPTE---KPTVPSSELYTIPTEKPMVHMEKPIVHT--EKPTVPT-EKP 365
 QY 466 APTTPKAAAPNTPKERA---PTTPKEAPPTTPKEAPPTTPKEAPPTTPKGAAPTTLKEP 521
 DB 366 TTPTEKSTVPT--KKPTVFKETLPE-GPTVPAL-RPTTPREGPAVPPKG--PTVLTE- 418
 QY 522 APTTPKKAAPKLAPTTPTEKSTSDKAPPTTPKGAAPT--TPKEAPPTTPKEAPPTTPK 580
 DB 419 -----WPTSHTEKSTVTEKILPTGKSTIPTEKPMVPTPKRI-----TTP- 458
 QY 581 GTAPPTLKEAPPTTPKKAAPKLAAPT--TTKGPTSTT---SOKPAPPTTPKEAPPTTPKEPA 636

DB 459 -TEKTIIPAKETPV---PIEKPMVPTERTTIPTERTTIPTERTIPTERTIPTERTIPTERTI 512
 QY 637 PTPPKKAPATTPETPP-----TTSVSTPT-----TTKEPTTIKSPDESPTPELSAEP 686
 DB 513 VPTEKIVPTTEKHTIPTERTIYLTERTTIPTERTIPTERTIPTERTIPTERTIPTERTIPT 571
 QY 687 PKALENSPREPGVPTTKTAAATKPEMTTAKDTERDRTTPEYTTAAPKMTKETATTT 746
 DB 572 -----EPTIPTKLVPT--ERTTTPKRTTPTIRITTTPTI-----RTTPT 613
 QY 747 EKTTESKITATTVQVSTTTQDTTPKTIITLKTTLIAPVTTTKTITTTTEIMNPEETA 806
 DB 614 ERTT-----TPIRTTPTERTT--IPTKKT-----VPEKTIIPPT-----ERTI 652
 QY 807 KPKDRATNSKATTPKQKTKAKPKPTSKKKPKTTPRVKRPKPTTPRKMTSTMPELNP 866
 DB 653 AP-----TTPDP-SPTLVPTOPAAVVPST-----SATTVPTRTTIASCP--FN 693
 QY 867 SRIAEAMLOTTTRPNOTPNKL 888
 DB 694 APERCACPVSQ-SPTPNCEL 714

Search completed: April 26, 2002, 16:16:35
 Job time: 186 sec

STATE: Virginia
COUNTRY: United States
ZIP: 2213-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/479,537A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 90/13101
FILING DATE: 23-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR91/00835
FILING DATE: 23-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/039,320
FILING DATE: 04-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,576
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 017753-025
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2035 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Peptide
LOCATION: 128..1899
OTHER INFORMATION: /note= "The amino acids spanning
OTHER INFORMATION: 128 to 1899 constitute a repeated region wherein the repeat
OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such
OTHER INFORMATION: repeats varies from 1 to 40."
FEATURE:
NAME/KEY: Peptide
LOCATION: 134
OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa
OTHER INFORMATION: Xaa Xaa which is the codon for Pro or Ala wherein Pro = CCT,
OTHER INFORMATION: CCC, CCA, or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 144
OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa
OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA
OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
FEATURE:
NAME/KEY: Peptide
LOCATION: 147
OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..21
OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
OTHER INFORMATION: 21 amino acid precursor sequence."
US-08-479-537A-2

Query Match 8.5%; Score 423.5; DB 2; Length 2035;
Best Local Similarity 24.4%; Pred. No. 3.9e-19;

Matches	235; Conservative	63; Mismatches	461; Indels	203; Gaps	48;
QY	75 TSLTVNKETVEETKETTITTKKQSTOCKETTSKAKETQSTIEKKSAND-LAPNSKVLAKPP	133			
Db	16 TVLTV-----VYSGHASTPGEKETTSATQSRSSVPSSTENAVSMISSVLSHS	65			
QY	134 P-KAETTKGP--ALTPKEP-----PTTPKEPASTTKKEPPTTIKSAP	176			
Db	66 PGSGSSITGQODVTLAPATPAGSNAATWGQDVTSVPTPALGSTTPPAHDVT---SAP	122			
QY	177 TTPKEPAPTT-----KSAPTTPKP-----APTTKEPATTTKKEPATT	217			
Db	123 --DNKPAPGSTAPXAHGVTAPDXRPXGSTAPXAHGVTAPDXRPXGSTAPXAHGVT	180			
QY	218 TKE--PAPTT-----KSAPTTPKPAPTTPK--KSAPTTPKPAPTTKEPPTT	264			
Db	181 APDXRPXGSTAPXAHGVTAPDXRPXGSTAPXAHGVTAPDXRPXGSTAPXAHGVT	240			
QY	265 PKE--PAPTTKEP-----APTTPKPAPTPAK--KSAPTTKEPAPTTKEPAPTT	311			
Db	241 APDXRPXGSTAPXAHGVTAPDXRPXGSTAPXAHGVTAPDXRPXGSTAPXAHGVT	300			
QY	312 TKEPSPPTPKPAPTT--TSAPTTKEPAPTT-----TSAPTTKEPSPPTTK--E	360			
Db	301 APDXRP--XPGSTAPXAHGVTAPDXRPXGSTAPXAHGVTAPDXRPXGSTAPXAHGVT	359			
QY	361 PAPTPKEPAPTPPK-----KPAPTTKEPAPTPPK-----EPAPTTTKKPAATPKEPAPT	412			
Db	360 SAPXRPXPGSTAPXAHGVTAPDXRPXPGSTAPXAHGVTAPDXRPXPGSTAPXAHGVT	419			
QY	413 TPKEPAP-----TPPKLTPPTPEKLAFTTPPKPAFTTPEELAPTTPEEPPTTPEEAPT	468			
Db	420 SAPXRPXPGSTAPXAHGVT-----APDXRPXPGSTAPXAHGVTAPDXRPX--PGST	471			
QY	469 TPKA-----AANTPKPAPPTPKPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPT	523			
Db	472 APXAHGVTAPDXRPXPGSTAPXAHGVTAPDXRP--XPGSTAPXAHGVTAPDXRPXPGS	530			
QY	524 TTP-----KKPAPKELAPTTTKEPTSTSDKPAP--TTPKGAPTTPKPAPPTT	571			
Db	531 TAPXAHGVTAPDXRPXPGSTAP--XAHGVTAPDXRPXPGSTAPXAHGVT-----APDXR	585			
QY	572 KEAPATTGK-----TAPTTTKEPAPTT-----KKPAPKELAPTTTKEPTSTSD	617			
Db	586 PXPGSTAPXAHGVTAPDXRPXPGSTAPXAHGVTAPDXRPXPGSTAP--XAHGVTAPDX	644			
QY	618 KPAP--TTPK-----ETAPTTKEPAPTT-----KKPAPTTPEPTTSEVSTP	661			
Db	645 RPXPGSTAPXAHGVTAPDXRPXPGSTAPXAHGVTAPDXRPXPGS--TAPXAHGVTAP	702			
QY	662 TTTKEPPT-----IKSPDESTPELSAEPTPKALENSPEKPGVPTTKTAA-----	707			
Db	703 DXRPXPGSTAPXAHGVTAPDXRPXPGSTAPXAHGVTAPDXRPXPGSTAPXAHGVTAP	762			
QY	708 -TKPEMTTAKDK---TTERDLRTPEPTTAPAKKTKETATTEKTSKIATTTQVTS	763			
Db	763 DXRPXPGSTAPXAHGVTAPDXRPXPGST--APXAHGVTAPDXRPXPGSTAPXAHGVT	820			
QY	764 TTTQDTPEFKITTTLTTLAPKVTITTKTITTTTEIMNKPETAKPKDRAVNSKATTPRQ	823			
Db	821 APDXRPXG-----STAPXAHGVT-----SAPDXRPXGSTAPXAHGVTAPDXRPX	868			
QY	824 KPTKAPKPKETSTKKKTPMRVKKRPTTTPPKMTSTMELNPT--SRLEAMLQTTTRNQ	882			
Db	869 -GSTAPXAHGVTAPDXRP--XPGSTAPXAHGVTSA-PDXRPXPGSTAPXAHGVTAPDX	924			
QY	883 TP 884				
Db	925 RP 926				

RESULT 14
US-09-083-116-2

Query Match	8.5%;	Score 423.5;	DB 4;	Length 1867;
Best Local Similarity	24.4%;	Pred. No. 3.5e-19;		
Matches 235;	Conservative 63;	Mismatches 461;	Indels 203;	Gaps 48;

[illegible]

RESULT 13
US-08-479-537A-2
Sequence 2, Application US/08479537A
Patent No. 5861381
GENERAL INFORMATION:
APPLICANT: CHAMON, Pierre
APPLICANT: KIENY, Marie-Paule
APPLICANT: LATHE, Richard
APPLICANT: HAREUVEN, Marc
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
TITLE OF INVENTION: 5
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATTHEWS, L.L.P.
STREET: P.O. Box 1404
CITY: Alexandria

TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 836-6620
 TELEFAX: (703) 836-2021
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1867 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 FEATURE:
 NAME/KEY: Peptide
 LOCATION: 128..1727
 OTHER INFORMATION: /note="The amino acids spanning 128 to 1727 constitute a repeated region wherein the repeat
 OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such
 OTHER INFORMATION: repeats varies from 1 to 40."
 FEATURE:
 NAME/KEY: Peptide
 LOCATION: 134
 OTHER INFORMATION: /note="Amino acid 134 is X1 = Xaa
 OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
 OTHER INFORMATION: or CCG, and Ala = GCT, GCC, GCA, or GCG."
 FEATURE:
 NAME/KEY: Peptide
 LOCATION: 144
 OTHER INFORMATION: /note="Amino acid 144 is Y = Xaa
 OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA
 OTHER INFORMATION: or ACG, and Asn = AAT or AAC."
 FEATURE:
 NAME/KEY: Peptide
 LOCATION: 147
 OTHER INFORMATION: /note="Amino acid 147 is X2 = Xaa
 OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
 OTHER INFORMATION: or CCG, and Ala = GCT, GCC, GCA, or GCG."
 FEATURE:
 NAME/KEY: Peptide
 LOCATION: 1..21
 OTHER INFORMATION: /note="Amino acids 1 to 21 are a
 OTHER INFORMATION: 21 amino acid precursor sequence."
 US-08-479-537A-5

Query Match 8.5%; Score 423.5; DB 2; Length 1867;
 Best Local Similarity 24.4%; Pred. No. 3.5e-19;
 Matches 235; Conservative 63; Mismatches 461; Indels 203; Gaps 48;
 QY 75 TSLVNETVETKETTITTKOTSDGKERTTSKETSIEKTSKAD-LAPTSKVLAKPT 133
 DB 16 TVLVV-----VTGSGHASTPGKEKETSATQSSVSPSTKKNVAVSWTSSLSHS 65
 QY 134 P-KAETTKGP--ALTTPKEP-----TPTTPKEPSTPKPEPTPTTKISAP 176
 DB 66 PGSSSTTGGQDVTLAPATEPAGSSAATGQDVTSVPTPALGSTITPPAHADV---SAP 122
 QY 177 TTPKEPAPTTT-----KSAPTTPKPEP-----APTTTKPEAPTTPKPEAPTT 217
 DB 123 --DKKRAPGSAAPXAHGVTSAPOKRPXGSAAPXAHGVTSAPOKRPXGSAAPXAHGVT 180
 QY 218 TKE--PAPTTT-----KSAPTTPKPEPAPTTK-----KRAPTTPKPEAPTTKEPTPTT 264
 DB 181 APDXRPXGSAAPXAHGVTSAPOKRPXGSAAPXAHGVTSAPOKRPXGSAAPXAHGVT 240
 QY 265 PKE--PAPTTKEP-----APTTPKPEPAPTPK-----KRAPTTPKPEPAPTTPKPEAPTT 311
 DB 241 APDXRPXGSAAPXAHGVTSAPOKRPXGSAAPXAHGVTSAPOKRPXGSAAPXAHGVT 300
 QY 312 TKESPPTPKPEPAPTT--TKSAPTTTKPEPAPTT-----TKSAPTTPKPEPSPPTTK----E 360
 DB 301 APDXRP--XPGSTAPXAHGVTSAPOKRPXGSAAPXAHGVTSAPOKRPXGSAAPXAHGVT 359
 QY 361 PAPTTPKPEPAPTTK-----KRAPTTPKPEPAPTTK-----EAPTTTKKRAPTAPEPAPTT 412

DB 360 SAPDXRPXGSAAPXAHGVTSAPOKRPXGSAAPXAHGVTSAPOKRPXGSAAPXAHGVT 419
 QY 413 TPKETAP-----TTPKKLPTTPEKLAPTTPEKAPTTTPEELAPTTPEPTPTTPEAPPT 468
 DB 420 SAPDXRPXGSAAPXAHGVTSAPOKRPXGSAAPXAHGVTSAPOKRPXGSAAPXAHGVT 471
 QY 469 TPKA-----AAPNTPKPEPAPTTPEKAPTTTPEKAPTTTPEKAPTTTPEKAPTTTPEKAP 523
 DB 472 APXAHGVTSAPOKRPXGSAAPXAHGVTSAPOKRPXGSAAPXAHGVTSAPOKRPXGSAAPXAHGVT 530
 QY 524 TTP-----KKRAPKELAPTTTKEPTSTTSOKPAP--TTPKGTAPTTTKEPAPTTT 571
 DB 531 TAPXAHGVTSAPOKRPXGSAAPXAHGVTSAPOKRPXGSAAPXAHGVTSAPOKRPXGSAAPXAHGVT 585
 QY 572 KEAPPTTPG-----TAPTTLKKEPAPTT-----KKRAPKELAPTTTGTSTSD 617
 DB 586 PEXGSAAPXAHGVTSAPOKRPXGSAAPXAHGVTSAPOKRPXGSAAPXAHGVTSAPOKRPXGSAAPXAHGVT 644
 QY 618 KRAP--TTPK-----ETAPTTPKPEPAPTT-----KKRAPTTPEPTTSEVSTP 661
 DB 645 RPXGSAAPXAHGVTSAPOKRPXGSAAPXAHGVTSAPOKRPXGSAAPXAHGVTSAPOKRPXGSAAPXAHGVT 702
 QY 662 TTPKEPTT-----TKSPDESTPELSAPPTPKALENSPKKEGVTPTTTPAA----- 707
 DB 703 DXRPXGSAAPXAHGVTSAPOKRPXGSAAPXAHGVTSAPOKRPXGSAAPXAHGVTSAPOKRPXGSAAPXAHGVT 762
 QY 708 --TKPEMTTAKDK---TTERDLRTPEPTTAAAPKMTKETATTEKTESKITATTTOVTS 763
 DB 763 DXRPXGSAAPXAHGVTSAPOKRPXGSAAPXAHGVTSAPOKRPXGSAAPXAHGVTSAPOKRPXGSAAPXAHGVT 820
 QY 764 TTQDPTPEKITTLLKTTTLAPKVTITTKITTTTEINNKKEEETAKPKPRATNSKATTPKQ 823
 DB 821 APDXRPXG-----STAPXAHGVT-----SAPDXRPXGSAAPXAHGVTSAPOKRPXGSAAPXAHGVT 868
 QY 824 KPKKAKKPKTSTKPKTTPVRKPKTTPPKMTSTMPELNPT--SRIEAMLOTETTRPQ 882
 DB 869 -GSTAPXAHGVTSAPOKRP--XPGSTAPXAHGVTSA--POKRPXGSAAPXAHGVTSAPOKRPXGSAAPXAHGVT 924
 QY 883 TP 884
 DB 925 RP 926

RESULT 12
 US-09-083-116-5
 ; Sequence 5, Application US/09083116
 ; Patent No. 6203795
 ; GENERAL INFORMATION:
 ; APPLICANT: CHAMBOY, Pierre
 ; APPLICANT: KIENEY, Marie-Paule
 ; APPLICANT: LARHE, Richard
 ; APPLICANT: HAREUVENT, Mara
 ; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
 ; TITILE OF INVENTION: TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
 ; NUMBER OF SEQUENCES: 5
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
 ; STREET: P.O. Box 1404
 ; CITY: Alexandria
 ; STATE: Virginia
 ; COUNTRY: United States
 ; ZIP: 22313-1404
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; FILING DATE: US/09/083,116
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/479,537

```

: RESULT 10
: Sequence 5, Application US/09219849
: Patent No. 6150081
: GENERAL INFORMATION:
: APPLICANT: VAN HEERDE, GEORGE V.
: APPLICANT: VAN RIJN, ALEXIS C.
: APPLICANT: BOUWSTRA, JAN B.
: APPLICANT: DE WOLF, FREDERIK A.
: APPLICANT: MOOBROEK, ANDREAS
: APPLICANT: WERTEN, MARC W.T.
: APPLICANT: WIND, RICHIELE D.
: APPLICANT: VAN DEN BOSCH, TANJA J.
: TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
: TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
: FILE REFERENCE: 2728-2
: PREPARATION THEREOF
: CURRENT APPLICATION NUMBER: US/09/219,849
: CURRENT FILING DATE: 1998-12-23
: NUMBER OF SEQ. ID NOS: 50
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 5
: LENGTH: 960
: TYPE: PR1
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence: Illustrative
: OTHER INFORMATION: amino acid sequence
: US-09-219-849-5

```

```

Query Match Similarity 9.0%; Score 452; DB 4; Length 960;
Best Local Similarity 25.6%; Pred. No. 2.8e-21;
Matches 173; Conservative 60; Mismatches 309; Indels 134; Gaps 41.

QY 132 PTPKAEATTTKGPAALTPPKKEPTPTTKKEPASTPTTKKEPTPTTKSAPTTKKEPAP--TTTKS 189
      :::::
Db 72 PEPAPGPPGSRDGP--PGGAPGAGP--PGSRDPGPGAGG--PAGPPGSRDPPPPAGPFA 126
      :::::

QY 190 APTTKKEPAPTTTKKEPA--TTTKKEPAPTTTKKPAPTTTTSAPTTTKKEPAPTT--KKPA 245
      :::::
Db 127 GPPGSRDGPFPAPPPAPGPPGSRDGPFGAPGPPGAPGPPGSRDGPAGPAPGSRDGP 186

QY 246 PTTTKKEPAPTTT--KEPTPTTKKEPA--PTTKKEPAPTTTKKEPA----- 284
      :::::
Db 167 P--PGAPGPAGGPPGSRDGPFGAPGAPGPPGSRDGPFGAGHPAGPKGAHGPAGKGAHG 244

QY 285 -----PTAPK-KKAPPTT--KEBPAPTTTKKEPA--PTTKKESSPTTKKEPAPTTTKS 330
      :::::
Db 245 PAGPKGAHGPAGPKGAPPPAGGPPGSRDGPFGAPGPPGAPGPPGSRDGP--PGAPGP-----A 298

QY 331 APTTKKEPAPTTTKSAPTTTKKEPSSTTTTKKEPAPTTTKKEPAPTTT--KKAPPTTKKEPAP 387
      :::::
Db 299 GPSPGSRDGP-----PGAPGPAGGPPGSRDGP--PGAPGPAGGPPGSRDGP--PGAGGP 348

```

```

0Y 388 TTP---KEAPRTTKKP-----PNAKEPAPPTPKETAPTPPKLTPTTPKLAPTT 433
Db 349 ACPGPGSRDGPBGAPGAPGPGSGRDPGPGAPGAPGPGSGRDP---PGAPGAPPG 404
0Y 438 PEKPAPT-TPEELATTEEPPTTPEEPATPTTAA--PNTKEPAPTTP---KEAPRT 493
Db 405 SRDPGPGAPGAPGPGSGRDPG--PGAPGAPGPGSGRDPGAPGAPGAPGSGRDPG 463
0Y 493 TKELA-----PTPKET-APTPPKT-----APTLKEPAPTPPKP-APKELA 535
Db 463 GAHGAPGKGAHGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPG 522
0Y 536 PTTKEPTSTSDKAPAPTPPKGA-PTTPKRAPTTT---KEAPTPKGA-PTTLKEP 590
Db 523 PGSGRDPGPGCA--PGAPGPGSGRDPGPGAPGAPGPGSGRDPGAPGAPGPGSGRDP 580
0Y 591 APTPPKKPA-----KELATTTKGPST--TSDKAPAPTPPKETATPTTKAPAPTPPK 642
Db 581 GP--PGAPGAPGPGSGRDPGPGAPGAPGAPGPGSGRDPGAPGAPGAPGSGRDPG--PGA 636
0Y 643 PAPTP---ETPPTTSEVSTPTTKKEPTTIHKSPESTPELSAEPTPKALEN-----S 693
Db 637 PGAPGPGSGRDPGPGAPGAPGAPGSGRDPG--PGAPGAPGPGSGRDPGPGAPGAPG 695
0Y 694 PKEGPVPTTKPAATK 709
Db 696 SKDPGPGAHGAPGPK 711

```

RESULT 1133
US-08-479-537A-5
Sequence 5, Application US/08479537A
Patent No. 5861381
GENERAL INFORMATION:
APPLICANT: CHAMBON, Pierre
APPLICANT: KIENY, Marie-Paule
APPLICANT: LARHE, Richard
APPLICANT: HAREUVENT, Mera
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/479,537A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 90/13101
FILING DATE: 23-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR91/00835
FILING DATE: 23-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/039,320
FILING DATE: 04-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,576
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Rodin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 017753-025

Db 1133 DEAKKQADKGNKGLVLPNTNSINKDPVNTQYSNITGNINPEKGVIPGSLPGLNYP 1192
QY 675 DESTPELSAEPPIKALE---NSKEGCVPTTKPATKPEMTTADKTERDLRTTPEP 731
Db 1193 SFMTPOOTBEITKPKVDYVGLPYDSTGEIIPDAKRLPGSVADDELITTEVLTITBE 1252
QY 732 TTAPAKMTKETATTEKTESKITATTTQVSTTTQDTTPPKITLTKTTLAKRVTTTK 791
Db 1253 VTGLP-IDLETTGLPRD-----VSLGPLPGLNITVD-----PSNKK 1287
QY 792 TI-----TTTEIMNKPEETAKPD-----RATNKK 816
Db 1288 PIGSHSGFINGTSGECSHKDPSTGKPLDPNTGLHPFEDSGSLINPETHKLGSGSHG 1347
QY 817 AITPKPKP-----TKAPKPTSTKPKPTMPVARKPKTTPRKMTST-----859
Db 1348 TFMVPKPKPGENGIMTPPOILEALNKLPTSEVNISPR---PSSDAVDRPTNTMMNK 1404
QY 860 -----MPELNPSTRIAEAMLOTTTRPNQ-----PNSKL-----VEYNPKS 895
Db 1405 ISGOTYVDKKTILGSAASVITHALCTPTQDTPTGLPSDPTGLPIFGFVNLVDPQT 1464
QY 896 EDAGAGETPHMLLRPHVEMPEVTPMDY-LPRVNOGIIINPM 939
Db 1465 GEQ--IKGSVYVSL--YVKEKNIVTEAAYGLPVPKTPGPIDPT 1505

RESULT 9
US-08-928-361B-6
Sequence 6, Application US/08928361B
Patent No. 6071518

GENERAL INFORMATION:

APPLICANT: Petersen, Carolyn
TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
TITLE OF INVENTION: SPECIES INFECTIONS
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: PETERS, CARYN, JONES & BIKSA
STREET: 385 Sherman Avenue, Suite 6
CITY: Palo Alto

COUNTRY: USA

ZIP: 94306-1840

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08-928, 361B

FILING DATE: 12-SEP-1997

CLASSIFICATION:

PRIORITY APPLICATION DATA:

APPLICATION NUMBER: US 60/026,062

FILING DATE: 13-SEP-1996

ATTORNEY/AGENT INFORMATION:

NAME: VERNY, HANA

REGISTRATION NUMBER: 30,518

REFERENCE/DOCKET NUMBER: 480,76-1(HV)

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-324-1677

TELEFAX: 650-324-1678

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 1721 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-928-361B-6

Query Match 9.2%; Score 463.5; DB 3; Length 1721;
Best Local similarity 22.1%; Pred. No. 9.7e-22;
Matches 271; Conservative 78; Mismatches 467; Indels 413; Gaps 39;

5 KNNRKKRPPKPPVADAGSLGNDGPKVTPDSTTHNKNVSTSPKITTAKPINRPS 64
Db 116 KSNRKK---TTPPSANTYAGVRSN-ETKTEPSANT---NELVADPKI-----N 158
QY 65 LPPNSDTSKE-----TSLTVNKETVEKETTITTKKOSTGCKEKTSAK 109
Db 159 AHCNSENSEGQOIFDMGSKVYIPYTKCVGKHNTTTTTTTTTTTTTTTTTTTT 208
QY 110 EQSIEKTAKDLAPTSKVLAKPTPKAETTTGPAITTPKEPTTPKEAPSTPKPEPTP 169
Db 209 -----TTT 240
QY 170 TTKSAPTPPKPEAPPTTKSAPTPPKPEAPPTTKPEAPPTTKPEAPPTTKSA 229
Db 241 TT 300
QY 230 PTPPKPEAPTPPKKAPPTPKPEAPPTPKPEPTTPPKPEAPPTTKPEAPPTPK 289
Db 301 TTTTKKPTTT 350
QY 290 KPAETPKPEAPTPPKPEAPPTTKPEPTTPPKPEAPPTTKSAPTTKPEAPPTKSAPT 349
Db 351 -----TTTKKPTTTTATTTT 406
QY 350 PKR-----PSPTT-----357
Db 407 TSETESVIRKDEMCWLEKNGCEAKGATYGVIGKDGRIENGMAFTMINDDTHVFRKK 466
QY 358 -----TKPE-----APTTPKE 368
Db 467 VKDVGNTSVRCRKGAGKLEFPDRSLDFTTPPVAGHNSCSIIYVSGDKIHSPYSGSD 526
QY 369 PAPTTPPKPAPTTPKE-----PAPTTPKEAPPTTKKPAATKAPKA 410
Db 527 VSLIS---APIQSELFNBYVCDCDTAKYGAHSGVGSADVITTTAKPPTTT--TGA 580
QY 411 PTPKEAPTPPKKLPPTPEKLAPTPKPEAPPTPEKLAPTPEPTTPPEAPPTP 470
Db 581 PGQPTTTTGSPPSKPTTTTAKATTT-----TLLNPITTTTOKPPTTT--TT- 627
QY 471 KAAPNTPKPEAPTPPKPEAPPTPKPEAPPTPKETAAPTTPKGTAPTTPKAPPTPKKA 530
Db 628 --KVPGRPIATTTTILKPIVTTTTKATTTTTTTPVPTT-----TTTTKDEMTTTTTPL 680
QY 531 PK---ELAPTTKEPNS-----TSDKPAPTTPKG-TAPTTPKEAPPT 570
Db 681 PDIGDIEITPIPIKMDKYRMIDYNSGLLSDNSDEPIPGSAGQIADISNLFVQTH 740
QY 571 PKPEAPTPPKGTAPTTPKE---PAPTTPKPAPELAAPTTPKGTSTTSKPAAPTTPKET 627
Db 741 KSTGLPIDPMVGLPDRKSGNLVHPYNTQMSGLSVSLAAKNLFTVPTDERYG--LPIDT 798
QY 628 APTTPKEAPTPPKKAPAPTTPPEPTTPPTSEVSTPT-----TKPEPTIHK 673
Db 799 LTGYLDVSLIPFN--PETGELFDPIISDEIMNTIAGIVSGISASISLSSQKALIDPA 856
QY 674 -----PDEST-----BELSAEPTPKALENSPK 695
Db 857 TNNVVERFGLINPATGMIIFGLPSQOTOPSEIEHGGIILPEVAANADKFKLSIP- 915
QY 696 EPGVPTTKPATKPEMTTAKDKTTER-----DLTTPETTTAPAKM 738
Db 916 -PSVP-----ESIP-----KKQKIDISSELMYDIESGRILGOVSKRPIDGSIAGDLP 963
QY 739 TKEIATTEKTESKITAT--TQVSTTTQDTTPPKITLTKTTLLAPVTTTKKTIITTE 797
Db 964 IMKTPTQTDSTVYKPIDPTTGLPFPNPFGHLINPNNNTMDSFAGAKYAVANSIKIDN 1023

```

: ORGANISM: Cryptosporidium parvum
US-08-700-651-5

Query Match      9.3%; Score 466.5; DB 3; Length 1721;
Best Local Similarity 20.9%; Pred. No. 6.3e-22;
Matches 285; Conservative 100; Mismatches 473; Indels 507; Gaps 45;

QY   33 KVTTPDSTTQHNNKVSFSPKITTAAPKLINRPSLPNSDSKETSILVAKETTVKEKETT 92
    || :||| :| :|| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db   190 KHNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 249
QY   93 TNKTSTSDGEEKTSKAEKOSIEKTSAKDLAPTSKVLAKPPKAETTTTCGPALTTPKEPT 152
    | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db   250 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 309
QY   153 PTTYKEPASTTPPEKPTTIKSAPTTKPKAPATTTKSAPTPKREAPATTTPKE 212
    || :||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db   310 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 369
QY   213 PAPITTKEPATTTKSAPTPKREAPATTTPKKAPATTTPKREATTTPKE----- 267
    || :||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db   370 TTTTTTTTTTKPTTTTTTTTT---TTTKKPTTTATTTTTSSNESVIKDEMCLE 423
QY   268 -----PAPTYK----- 273
Db   424 KNCECAKATIVGVIGKDRLNGMAFTMLPNDTIHFRFKVKDVONTISVRCKAG 483
QY   274 -----EPATTPKBEA--TPAKKP----- 292
Db   484 KLEFPDRSIDFTIPVAGHNSCIIVGSGDGKIHVSPYSKDVLSIADIQSELENVEY 543
QY   293 -----PTTPKEAPATT---PKCAPATTKEPS--PTTPKEAP 325
Db   544 YCDTCAKIGAISHGYOXSADFVVTTTAAKPTTTTGTGARGPRTTTHGSPSKPTTTTGA 603
QY   326 TTTKSA----PTTKEPATTTKSAPTPKESPTTPAKAPATTTPKREAPATTPKRAP 380
    || :||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db   604 TTTTTTLNPIITTTQKRPTTTTTRKVPKRIATTTTTLKLPIVTTTTKATTTTTTTTVP 663
QY   381 TKREPAPTPPKRAPATTTKKAP-----TPAKE 408
Db   664 T-----TTTKRDEMTTTTPLPDIGIEITPPIEKMIDKYRMIVDYNSGLIDSNE 718
QY   409 PAP-----TPKETP--PTTP----- 422
Db   719 PLPGSQAGIADTSLNPLPVQTHKSTGLPIDBMGLPRDKSGNLVHPYTNOTMGLSVSY 778
QY   423 ---KLLT-----PTTPKEIAPTPBE----- 439
Db   779 LAAKMLIVTDIETGLPIDLTLYGRPLDPVSLIPFNETGETLPDISDELINGTACIYSG 838
QY   440 -----KPAPTPREE-----LAPTT-----PEEPTTPP--BEPAPT 468
Db   839 ISASESLISQSALIDPATNMVYGEEGGLLNPTAGVMIRGFLEPGEOTOFSPEDIGII 898
QY   469 TPKAAPNTPKREAPATTPKREAPATTPE----- 496
Db   899 PPEVAANAADKFKISIPSPVESISEPKDXIDISISLMDISSRLIGOVSKRPIGSA 958
QY   497 ----PAPTPKREAPATTPKSTAPTTTLKEAPATTPKKAPKELAPT--TTKEPTSTTSDP 550
Db   959 GDNLNPIMKTPQTIDISVGNKPIDPT---GLPKNP--PTGHLLNPTNNMTMOSSFAGAVKY 1013
QY   551 APPT-----PKSTAPTTPKBEA-----PTTPKRAPATTTPK----- 580
Db   1014 AVSNGIKTDWNYGLPVGIELGLEPKDGSDIPENSTTGELVDYSTCKPIINNSTAGIVSKP 1073
QY   581 GTAPT---TLKEAPATTPKKAPKELAP---TTTKGPTSTSDKAPATTTPKETAAPT 632
Db   1074 GLPFIEDENGNLDPSTNPLRIDGNQLVANPETHNSIVSGSTGTT--KPRGFLPVNGGVVP 1132

```


Matches	273;	Conservative	86;	Mismatches	459;	Indels	432.	Gaps	11.
---------	------	--------------	-----	------------	------	--------	------	------	-----

Qy	33	K	T	P	D	S	T	O	H	N	K	V	S	P	K	I	T	T	A	K	I	N	P	S	L	P	N	S	T	K	E	S	T	L	N	V	K	E	T	T	E	K	T	T		92										
Db	307	K	H	I	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T		366													
Oy	93	T	N	K	O	T	S	D	K	E	K	E	T	T	S	A	K	E	T	O	S	I	E	K	T	S	A	K	O	L	A	P	T	S	V	L	A	K	P	P	K	A	T	T	K	G	A	L	T	T	P	E	R	T		15
Db	367	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T		426									
Oy	153	P	T	T	P	K	E	A	S	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T		210									
Db	427	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T		468									
Oy	211	K	E	B	A	P	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T		259									
Db	487	T	T	T	A	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T		546									
Oy	260	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T		266									
Db	547	A	K	G	A	T	V	G	I	K	D	G	R	I	E	N	G	A	T	T	M	I	R	P	N	D	T	I	H	R	F	K	V	K	D	O	G	N	I	S	R	C	G	A	K	L	E	F	D		606					
Oy	267	E	P	A	P	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T		291									
Db	607	R	S	L	D	F	I	T	P	V	A	G	H	N	S	C	S	I	I	V	G	S	G	G	K	I	H	S	P	Y	S	K	D	V	S	L	I	S	A	I	O	C	E	L	F	N	E	Y	C	D	T		666			
Oy	292	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T		330									
Db	667	A	K	G	A	I	H	S	G	Y	O	T	S	A	D	E	V	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T		726									
Oy	331	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T		386									
Db	727	L	N	P	I	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T		781									
Oy	387	P	T</																																																					

```

QY 729 TKATATTEKTESKIATITTOYTSTTDOITPPKITTLLTKTLTAARVYTTAKITTYTE 797
    || || || || || || || || || || || || || || || || || || || || ||
Db 1376 DLEGTGPRDP-----VSGLPOLNGLTVL-----PSNKKPLPGSHS 1411
QY 798 --INMKPEETAKPKDRATNSKATTPKQKQ--TKAP--KKPTSKKKEKTPYRKAKT--- 849
    || || || || || || || || || || || || || || || || || || || || ||
Db 1412 GFINGTSGEOSHEDPSTG-----PFLDNLTGLPDEDSGLIMPETDKLQCSHSQT 1465
QY 850 TPTPRKMTSTWELNPTSRIRAEALQTTTPNPOTPSKLVENPK--SEDA 898
    || || || || || || || || || || || || || || || || || || || || ||
Db 1466 MPVVGKRGQGGNGGIMPEQILLEL-----NKLPTSNENVNLSIPRSDSA 1508

```

RESULT 5
US-07-638-431-2
; Sequence 2, Application US/07638431
; Patent No 5198535

GENERAL INFORMATION:
APPLICANT: Hoffman, Stephen L.
APPLICANT: Charoenvit, Yupin
APPLICANT: Hedstrom, Richard
APPLICANT: Khumsith, Srisin
APPLICANT: Rogers IV, William O.
TITLE OF INVENTION: Protective malaria sporozoite surface protein
TITLE OF INVENTION: Immunogen and gene
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS: 2

ADDRESS: A. David Speack
STREET: NMRC Building 1 T-12 National Naval
STREET: Bethesda
CITY: Bethesda
STATE: MD
COUNTRY: USA
ZIP: 20814-5044
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.24
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/638,431
FILING DATE: 19910110
CLASSIFICATION: 43

CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Sweeney, Aaron D

TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 295-6759

TELEFAX: (301) 295-4033
INFORMATION FOR SEQ ID NO: 2

SEQUENCE CHARACTERISTICS:
LENGTH: 826 amino acids

```

TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE:

```

MOLECOLE 1
US-07-638-431-2

Query Match	9.78;	Score 488.5;	DB 1;	Length 826;
Best Local Similarity	26.48;	Pred No 1	20-22	

Matches	153; Conservative	54; Mismatches	223; Indels	149; Gaps	34.
---------	-------------------	----------------	-------------	-----------	-----

168 PTTIKSAPITPKEP-PTTKSAPITPKEP-PTTKP---APITPKP-APITPKE 230

273 TPCKVRDCQIPPIPIPNKIPKPSNPPEEPVNPDPNPNNPNPNPNPNPNPNPN 332

221 РАПТТТКАРПТПКЕР-АРПТПКР-АРПТПКЕР-АРПТПКЕР-ТРПТПКЕРАРПТКЕРА 270

[illegible]

277 PTPPKPEAPATPK--KPAPTTPKEAPPTTPKEAPTTTTPKEAPTTTTSAPT 334

[illegible]

```

Db      421 ADESIDHLPHESDCGCOYLQCVHGOTIARPCGNLHFSPATQSCSPVTAQGVCECDSD 480
OY      382 ---KKEPAPTPPKKPPAPTTTKKCAPPAKPEPAPTTTKEPAITPPKLIPTTPEKLAAPT 438
Db      481 NQSTAPTAAPTAAPTAAPTAAPTAAPTA--APSTVVPAPTPAPTAAPVPTT---AIPt-- 533
OY      439 EKPAPTTPEELAPTTPEPTTPEEPAPTPPKAAAPNT--KPEAPTPPKPEAPTTKE 496
Db      534 --PAPTAAPTAAPTAAPDESPTTIVP-PIAAPTAPTAIVAPVPIPTV---SAPTAAPT 587
OY      497 PAPTPKETAPTPPKGTAPTTLKEPAPTPPKKPAKELAPTTTKEPTSTTSOKPAPTPPK 556
Db      568 AAPTAAPTAAPTTAVPEIPTVTSPTAAETTAAP--APNTT-----VTVP 632
OY      557 GTAPTPKEAPTPPKPEAPTPPKGTAPTTLKEPAPTPPKKPAKELAPTTTGGPTSTTS 616
Db      633 TAAPT- AAPAPNTTVVPTPAAPTAPPVAA-----AINTTAAPVTTS 677
OY      617 DKPAPTPKETAPTPPKPEAPTPPKK 643
Db      678 ---APATPDEDLIDP--PLPNPPIP 699

RESULT 4
US-08-928-361B-5
; Sequence 5, Application US/08928361B
; Patent No. 6071518
; GENERAL INFORMATION:
; APPLICANT: Petersen, Carolyn
; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
; TITLE OF INVENTION: SPECIES INFECTIONS
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PETERS, VERNY, JONES & BIKSA
; STREET: 385 Sherman Avenue, Suite 6
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-1840
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928.361B
; FILING DATE: 12-SEP-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/026,062
; FILING DATE: 13-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: VERNY, HANA
; REGISTRATION NUMBER: 30,518
; REFERENCE/DOCKET NUMBER: 480.76-1(HV)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-324-1677
; TELEFAX: 650-324-1678
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1837 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-928-361B-5

Query Match          9.8%; Score 493; DB 3; Length 1837;
Best Local Similarity 21.8%; Pred. No. 1.4e-23;

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:13:29 ; Search time 49.78 seconds
(without alignments)
425.384 Million cell updates/sec

Title: US-09-556-246-1_COPY_200_1140
Perfect score: 5011
Sequence: 1 VKDNKKNTKKKKPKPPV.....DMDYLPVNOGIINPMLS 941

Scoring table:
BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA: *
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep: *
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep: *
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep: *
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep: *
5: /cgn2_6/ptodata/2/1aa/PTCUTS.COMB.pep: *
6: /cgn2_6/ptodata/2/1aa/backfile1.pep: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	522	10.4	805	US-09-103-429A-4	Sequence 4, Appli
2	506.5	10.1	744	5202236-25	Patent No. 5202236
3	498.5	9.9	786	US-09-103-429A-3	Sequence 3, Appli
4	493	9.8	1837	US-08-928-361B-5	Sequence 5, Appli
5	488.5	9.7	826	US-07-638-431-2	Sequence 2, Appli
6	488.5	9.7	826	PCT-US92-00018-2	Sequence 2, Appli
7	476.5	9.5	652	5202236-13	Patent No. 5202236
8	466.5	9.3	1721	US-08-700-651-5	Sequence 5, Appli
9	463.5	9.2	1721	US-08-928-361B-6	Sequence 6, Appli
10	452	9.0	960	US-09-219-849-5	Sequence 5, Appli
11	423.5	8.5	1867	US-08-479-537A-5	Sequence 5, Appli
12	423.5	8.5	1867	US-09-083-116-5	Sequence 2, Appli
13	423.5	8.5	2035	US-08-479-537A-2	Sequence 2, Appli
14	423.5	8.5	2035	US-09-083-116-2	Sequence 2, Appli
15	419.5	8.4	1185	US-08-276-967-2	Sequence 2, Appli
16	419	8.4	1185	US-09-041-886-23	Sequence 23, Appli
17	417	8.3	829	US-08-642-255-132	Sequence 132, Appl
18	417	8.3	829	US-08-397-633A-53	Sequence 53, Appli
19	417	8.3	829	US-08-175-155-68	Sequence 68, Appli
20	417	8.3	837	US-08-477-509B-103	Sequence 103, App
21	417	8.3	837	US-08-642-255-101	Sequence 101, App
22	417	8.3	837	US-08-707-237A-75	Sequence 75, Appl
23	417	8.3	837	US-08-482-085B-103	Sequence 103, App
24	417	8.3	837	US-08-397-633A-50	Sequence 50, Appl
25	413.5	8.3	907	US-08-783-774-2	Sequence 2, Appli
26	413.5	8.3	907	PCT-US95-04611A-19	Sequence 19, Appli
27	404.5	8.1	408	US-07-609-716-65	Sequence 65, Appli

28	404.5	8.1	408	4	US-08-475-411A-65	Sequence 65, Appli
29	404.5	8.1	408	4	US-08-478-029A-65	Sequence 65, Appli
30	398.5	8.0	682	1	US-08-642-255-126	Sequence 126, Appl
31	398.5	8.0	682	1	US-08-397-633A-36	Sequence 36, Appli
32	381	7.6	1848	4	US-08-296-791-6	Sequence 6, Appli
33	381	7.6	1848	4	PCT-US95-10661A-6	Sequence 6, Appli
34	376	7.5	960	4	US-09-219-849-6	Sequence 6, Appli
35	374.5	7.5	1537	1	US-08-325-267A-2	Sequence 2, Appli
36	371	7.4	761	2	US-08-707-237A-84	Sequence 114, Appl
37	371	7.4	762	1	US-08-642-255-114	Sequence 26, Appli
38	371	7.4	762	1	US-08-397-633A-26	Sequence 62, Appli
39	370.5	7.4	1064	1	US-08-642-255-62	Sequence 4, Appli
40	369.5	7.4	1231	3	US-08-904-263A-4	Sequence 120, Appl
41	368	7.3	762	1	US-08-642-255-120	Sequence 31, Appli
42	367.5	7.3	762	1	US-08-397-633A-31	Sequence 28, Appli
43	367.5	7.3	1187	1	US-08-320-559-28	Sequence 28, Appli
44	367.5	7.3	1187	5	US-08-545-860D-28	Sequence 28, Appli
45	367.5	7.3	1187	5	PCT-US94-04496-28	Sequence 28, Appli

ALIGNMENTS

RESULT 1
US-09-103-429A-4
; Sequence 4, Application US/09103429A
; Patent No. 6187558
; GENERAL INFORMATION:
; APPLICANT: Granados, Robert R
; TITLE OF INVENTION: A No. 6187558el Invertebrate Intestinal Mucin
; TITLE OF INVENTION: A No. 6187558el Invertebrate Intestinal Mucin
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESSES:
; ADDRESSES: Brown, Pimisi & Michaels, P.C.
; STREET: 118 No. 6187558th Tlloga
; CITY: Ithaca
; STATE: NY
; COUNTRY: USA
; ZIP: 14850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/103,429A
; FILING DATE: 24-JUN-1998
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Michaels, Christopher A
; REGISTRATION NUMBER: 34,390
; REFERENCE/DOCKET NUMBER: BTI-39
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (607) 256-3628
; TELEFAX: (607) 256-2000
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 805 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Trichoplusia ni
; TISSUE TYPE: peritrophic membrane
; US-09-103-429A-4

Query Match 10.4%; Score 522; DB 4; Length 805;
Best Local Similarity 27.8%; Pred. No. 8.9e-26;


```
Db 701 KKEEA---KEKKAAPPEEETPAKLGVEAKPKKEADAKA-----KEPSKP 744
OY 570 TPKEPAPTTPKGTAPTTLKEPAPTTPKKAPKRELAPTTKGPTSTSDKPAPTTPKETAP 629
Db 745 SEKE---KPK-----KEVPAAPEKKDTKE-----EKTESKKREKPKMEAK 784
OY 630 TPKEPAPTTPKKPAPTTPPETPTTSEVSTPTTKKEPTTIHSPDE 676
Db 785 AKED-----KGLPQEPSKPKTEKAKEKSSSTDKDSQPSKAPED 824
```

Search completed: April 26, 2002, 16:19:03
Job time: 328 sec

RESULT 15
 NF-H RAT STANDARD: PRT: 831 AA.
 AC P16884: 063368;
 DT 01-AUG-1980 (Rel. 15, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE NEUROFILAMENT TRIPLET H PROTEIN (200 KDA NEUROFILAMENT PROTEIN)
 GN (NEUROFILAMENT HEAVY POLYPEPTIDE) (NF-H) (FRAGMENT).
 OS Rattus norvegicus (Rat).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 CC NCBI_TaxID=10116;
 RX MEDLINE-69065087; PubMed-3143606;
 RX Brien K.C., Robinson P.A., Wilson D., Anderson B.H.;
 RT "Partial sequence of the rat heavy neurofilament polypeptide (NF-H).
 RT Identification of putative phosphorylation sites.";
 RT FEBS Lett. 241:213-218(1988).
 RN [2]
 RP SEQUENCE OF 37-831 FROM N.A.
 RA MEDLINE-88309090; PubMed-2457365;
 RA Datsigny A., Pham-Dinh D., Rousset C., Felix J.M., Nussbaum J.L.,
 RA Jolles P.;
 RT "The large neurofilament subunit (NF-H) of the rat: cDNA cloning and
 RT in situ detection.";
 RT Blochem. Biophys. Res. Commun. 154:1099-1106(1988).
 RN [3]
 RP SEQUENCE OF 1-88 AND 243-313 FROM N.A.
 RA MEDLINE-87080760; PubMed-2878828;
 RA Robinson P.A., Wilson D., Anderson B.H.;
 RT "Isolation of a cDNA for the rat heavy neurofilament polypeptide
 RT (NF-H).";
 RT FEBS Lett. 209:203-205(1986).
 RN [4]
 RP SEQUENCE OF 318-831 FROM N.A.
 RA MEDLINE-89146647; PubMed-2928342;
 RA Lieberburg I., Spinner N., Snyder S., Anderson J., Goldhaber D.,
 RA Smulowitz M., Carroll Z., Emanuel B.S., Breiner J., Rubin L.;
 RT "Cloning of a cDNA encoding the rat high molecular weight
 RT neurofilament peptide (NF-H): developmental and tissue expression in
 RT the rat, and mapping of its human homologue to chromosomes 1 and
 RT 22.";
 RL Proc. Natl. Acad. Sci. U.S.A. 86:2463-2467(1989).
 CC -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
 CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
 CC NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT
 CC SUBSERVED BY THE TWO SMALLER NF PROTEINS.
 CC -1- PTH: THERE ARE A NUMBER OF REPEATS OF THE TRIPLET K-S-P, NFH IS
 CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
 CC THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
 CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
 CC OF AXONAL CALIBER.
 CC -1- PTH: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
 CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
 CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
 CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
 CC -1- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 783
 CC ONWARD AND IS LONGER DUE TO A FRAMESHIFT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb.slb.ch).

CC EMBL: M37227; AAA41693.1; ALT_FRAME.
 DR EMBL: X13804; CAA32038.1; ALT_FRAME.
 DR EMBL: M21964; AAA41695.1; -
 DR EMBL: J04517; AAA41692.1; -
 DR PIR: A30796; A30796.
 DR PIR: A25649; A25649.
 DR PIR: B25649; B25649.
 DR PIR: S02003; S02003.
 DR InterPro: IPR001664; IF.
 DR Pfam: PF00038; filament; 1.
 DR PROSITE: PS00226; IF; 1.
 KW Intermediate filament; Coiled coil; Neurone; Phosphorylation; Repeat.
 FT NON TER 1
 FT DOMAIN 1
 FT 276 641 51 X 3 AA TANDEM REPEATS OF K-S-P.
 FT CONFLICT 164 164 L -> I (IN REF. 2).
 FT CONFLICT 185 185 I -> S (IN REF. 2).
 FT CONFLICT 193 193 L -> T (IN REF. 2).
 FT CONFLICT 199 199 M -> T (IN REF. 2).
 FT CONFLICT 346 346 K -> N (IN REF. 1).
 FT CONFLICT 373 373 A -> V (IN REF. 1).
 FT CONFLICT 482 482 G -> E (IN REF. 2 AND 4).
 FT CONFLICT 485 485 P -> S (IN REF. 2).
 FT CONFLICT 570 571 RK -> KE (IN REF. 2 AND 4).
 FT CONFLICT 591 591 P -> T (IN REF. 2 AND 4).
 FT CONFLICT 727 727 A -> V (IN REF. 4).
 FT CONFLICT 757 759 AAP -> GST (IN REF. 4).
 FT CONFLICT 769 769 T -> L (IN REF. 2).
 FT CONFLICT 775 775 R -> P (IN REF. 2 AND 4).
 SQ SEQUENCE 831 AA; 89486 MW; 1B0973C3F13EF768 CRC64;

Query Match 9.9%; Score 497; DB 1; Length 831;
 Best Local Similarity 27.6%; Pred. No. 2.3e-13;
 Matches 195; Conservative 59; Mismatches 307; Indels 146; Gaps 35;
 37 PDFTSTQHNKVSSTPKITAKPINRPSLPPNSDTSKETSLYNKKETVETKETTNNKQ 96
 197 PSMST--HIKVKSEKIKVVE-----KSEKETVVEQVEIQTVEVTEED 242
 97 TSDGKCKTTSKAKETQSIETSKADLAPTSKVLAKPPKAKETTKGALTTPREP---T 152
 243 KEAQGEEREAEGEGEAATTS-----PRAEASP-----EKETKSPVKEAKSPAEKS 293
 153 PTPKKEPA-STTPKEPTPTTIKSAPTTPKEPA-----PTTKSAPTTPKEPA-----PTTK 203
 294 PAAKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEK 352
 204 EP-----APTPKKEPA-----PTTKKEPA-----PTTKSAPTTPKEPA-----PTPKKPA-- 245
 353 SPGEAKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEK 411
 246 --PTTPKEPA-----PTTPKEP-----PTTPKEAPPTTPKEP-----PTTPKEPA-----PTA 287
 412 KSPVEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEK 470
 288 PKKPAPTPKEPAPTPKKEPAPTPKKEPAPTPKKEPAPTPKKEPAPTPKKEPAPTPKKEPAPTP 338
 471 AKSPADA--KSPAGAKPRAEAKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSP 528
 339 APTTTS-----APTPKKEPPTTPKEPA-----PTTKKEPAPTPPKKAPPTTPKKEPAPT 389
 529 VEGAKSLAEAKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEK 588
 390 PKKPAPTPKAPAPKAPAPKAPAPTPKKEPAPTPKKEPAPTPKKEPAPTPKKEPAPTPKKEPAPTP 449
 589 AKPRAEAKKRRADIRSPQVSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEK 644
 450 APTPEEPPTTPKEPAPTPKRAAPNTPKKEPAPTPKKEPAPTPKKEPAPTPKKEPAPT 509
 645 ---KAKEPKVVEEKTPATPKTEVKESKDEAPKEAQNPKA-EKKEPTLTKPKSPGKA 700
 510 PKGTAPTTLKEKAPPTPKKPAKELAPTTTKETSTTSKPAPTPKGAPPTPKKEPAPT 569

[illegible]

RESULT	14
RPBL_MOUSE	
ID	RPBL_MOUSE
AC	P08775;
DT	01-NOV-1988 (Rel. 09, Created)
DT	01-MAR-1992 (Rel. 21, Last sequence update)
DT	20-AUG-2001 (Rel. 40, Last annotation update)
DE	DNA-DIRECTED RNA POLYMERASE II LARGEST SUBUNIT (EC 2.7.7.6) (RPBL).
CN	POLR2A OR RP02.1 OR RP11215.
OS	Mus musculus (Mouse).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX	NCBI_taxid=10090;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=87280135; PubMed=3038894;
RA	Ahearn J.M., Jr., Bartolomei M.S., West M.L., Cisek L.J., Corden J.L.,
RT	"Cloning and sequence analysis of the mouse genomic locus encoding
RT	the largest subunit of RNA polymerase II.";
RL	J. Biol. Chem. 262:10695-10705(1987).
RN	[2]
RP	SEQUENCE OF 1587-1970 FROM N.A.
RX	MEDLINE=86068017; PubMed=2999785;
RA	Corden J.L., Cadena D.L., Ahearn J.M., Jr., Dahmus M.E.;
RT	"A unique structure at the carboxyl terminus of the largest subunit
RT	of eukaryotic RNA polymerase II.";
RL	Proc. Natl. Acad. Sci. U.S.A. 82:7934-7938(1985).
RN	[3]
RP	REVISIONS, AND PRESENCE OF AN ADDITIONAL EXON.
RX	MEDLINE=92176992; PubMed=1542581;
RA	Wintzertich M., Acker J., Vicaire S., Vigneron M., Kedinger C.;
RT	"Complete sequence of the human RNA polymerase II largest subunit.";
RT	Nucleic Acids Res. 20:910-910(1992).
CC	-1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC	OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC	SUBSTRATES.
CC	-1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N PYROPHOSPHATE +
CC	RNA(N).
CC	-1- SUBUNIT: RNA POLYMERASE II CONSISTS OF 12 DIFFERENT SUBUNITS.
CC	-1- SUBCELLULAR LOCATION: NUCLEAR.
CC	-1- PTM: THE TANDEM 7 RESIDUES REPEATS CAN BE HIGHLY PHOSPHORYLATED.
CC	THE PHOSPHORYLATION ACTIVATES POL2.
CC	-1- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE
CC	FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA
CC	PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE

CC	IITFOR 5S AND TRNA GENES.
CC	-I- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/ or_send_email_to_license@isb-sib.ch).
CC	-----
DR	EMBL; M12130; AAAA0071.1; '-'
DR	EMBL; M14101; AAAA0071.1; JOINED.
DR	PIR; A28490; A28490.
DR	MGI; G6198086; Rpo2-1.
DR	InterPro; IPR000684; RNA_polII_repeat.
DR	InterPro; IPR000722; RNA_pol_A.
DR	InterPro; IPR002879; RNA_pol_A2.
DR	Pfam; PF00623; RNA_pol_A; 1.
DR	Pfam; PF01854; RNA_pol_A2; 1.
DR	PROSITE; PS00115; RNA_POL_II_REPEAT; 42.
KM	Transferrase; DNA-directed RNA polymerase; Transcription; Zinc; Repeat;
KW	DNA-binding; Nuclear protein; Phosphorylation; Zinc-finger.
KM	C2H2-TYPE (POTENTIAL)
FT	ZN_FING 71 87
FT	DOMAIN 1550 1958
FT	CARBOXYL-TERMINAL 7-RESIDUE REPEATS.
FT	P -> R (IN REF. 1 AND 2).
FT	CONFLICT 1498 1498
FT	MISSING (IN REF. 1 AND 2).
FT	CONFLICT 1499 1536
SQ	SEQUENCE 1970 AA; 217175 MW; 7D76F38BD92AA657E CRC64;

Query Match	9.9%;	Score 498.5;	DB 1,	Length 1970;
Best Local Similarity	33.1%;	Pred. NO. 3.9e-13;		
Matches 180;	Conservative 101;	Mismatches 168;	Indels 95;	Gaps 47;

[illegible]


```

Db 117 PVAATPPVVGQIPVAAPVIAATPPVAASAPTAATVPVIAISPVAAPVPAATVTPVAPV 176
Oy 280 PKBA--PLAKKRAPPTPKKAPPTTKKAPPTTKK--PSPTTKKBPAP-----TTTKSA 331
Db 177 AAVPAAPVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAV 236
Oy 332 PTTTKKBPAPPTTKKAPPTPKKAPPTTKKBPAPPTTKKBPAPPTTKKBPAPPTTKKBPAP 387
Db 237 PLAAAPVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAV 296
Oy 388 TTP-----KEPAPPTTKKBPAPPTTKKBPAPPTTKKBPAPPTTKKBPAPPTTKKBPAP 428
Db 297 ETPLAPVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAV 356
Oy 429 TPKL-APTPKAPPTPKKAPPTTKKBPAPPTTKKBPAPPTTKKBPAPPTTKKBPAPPTTKK 484
Db 357 AAESIPAPVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAV 413
Oy 485 TPKAPPTTKKBPAPPTTKKBPAPPTTKKBPAPPTTKKBPAPPTTKKBPAPPTTKKBPAP 533
Db 414 V----LPPVPAAPVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPA 469
Oy 534 LAPPTTKKBPAPPTTKKBPAPPTTKKBPAPPTTKKBPAPPTTKKBPAPPTTKKBPAPPTTK 585
Db 470 AAPVSTTP--TTASVETTPAPPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAV 527
Oy 586 TLKBPAPPTTKKBPAPPTTKKBPAPPTTKKBPAPPTTKKBPAPPTTKKBPAPPTTKKBPAP 631
Db 528 APPEAAADLIEVEPPAPPTTKKBPAPPTTKKBPAPPTTKKBPAPPTTKKBPAPPTTKKBPAP 585
Oy 632 PKKAPPTTKKBPAPPTTKKBPAPPTTKKBPAPPTTKKBPAPPTTKKBPAPPTTKKBPAPPTTK 681
Db 586 EVAAPVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAV 641
Oy 682 SAEPTRKALNSKPEKGPVPTTKKBPAPPTTKKBPAPPTTKKBPAPPTTKKBPAPPTTKKBPAP 739
Db 642 TVE-PFEAAVAKLDPAT--TEAPVTTQEDVANINDGAPATET--TTPAVEITVTAAEVS 697
Oy 740 KEATTTKTESKITATTTQVST-----TTQDTPPKITTLKTTTLA-PKVTTTKKT 792
Db 698 DIAIPVADIPVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAV 757
Oy 793 ITTTELMNKEETAKKORATSKATTPPKOKPKAPKPKPKPKPKPKPKPKPKPKPKPKPKPKPK 852
Db 758 EAPITAGDPDNTSGVISEVPTIAKPEVEVPTS--EIEQSSSPS--DSVPVAKITPL 813
Oy 853 PKKMTMBELNPTSRIAEAM--LQTTTRPNOTPNSKLVEN-----PKS 895
Db 814 LRLDLOTTDVSLAIAATLDAIGEKLDKARNQVMDRLCEIKILGPKS 864

RESULT 11
FPL_MYTGO STANDARD: PRT: 872 AA.
AC Q25434;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, last sequence update)
DT 20-AUG-2001 (Rel. 40, last annotation update)
DE ADHESIVE PLAQUE MATRIX PROTEIN PRECURSOR (FOOT PROTEIN 1) (MCPFL).
CN FPL.
OS Mytilus coruscus (Sea mussel).
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloida;
OC Mytilidae; Mytilus.
OX NCBI_TaxID=42192;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Foot;
RX MEDLINE=96394686; PubMed=8798340;
RA Inoue K., Takeuchi Y., Takeyama S., Yamaha E., Yamazaki F., Odo S.,
RT "Adhesive protein cDNA sequence of the mussel Mytilus coruscus and
its evolutionary implications.";

```

```

RL 101. Evol. 43:348-356(1996).
CC -1- FUNCTION: PROVIDES ADHESIVENESS TO THE MUSSEL'S FOOT. MUSSELS
CC PRODUCE ONE OF THE STRONGEST WATER INSOLUBLE GLUE. THE MUSSEL'S
CC ADHESIVE IS A BUNDLE OF THREADS, CALLED A BYSSUS, FORMED BY A
CC FIBROUS COLLAGENOUS CORE COATED WITH ADHESIVE PROTEINS.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- TISSUE SPECIFICITY: PRODUCED BY THE BYSSAL GLAND.
CC -1- DOMAIN: ALMOST EXCLUSIVELY COMPOSED OF REPEATS OF A DECAPEPTIDE.
CC -1- PTM: PROLINES IN THE REPEAT ARE HYDROXYLATED (SINGLE OR DI-) AND
CC ALSO TYROSINE (THUS PRODUCING DOPA = 3,4-DIHYDROXYPHENYLAMINE).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb-sib.ch).
CC
DR EMBL: D63777; BAA09850.1; -
DR InterPro: IPR002964; Adhesive_plaq.
DR InterPro: IPR002965; P-rich_extensn.
DR PRINTS: PR01216; ADHESIVEI.
DR PRINTS: PR01217; PRICHEXTEN.
KW Signal; Repeat; Hydroxylation.
FT SIGNAL 1 20
FT CHAIN 21 872
FT DOMAIN 21 41
FT DOMAIN 124 872
FT FT
FT FT
FT DOMAIN 184 192
FT DOMAIN 213
FT SEQUENCE 872 AA; 101677 MW; 98CC70D7C75FEFC4 CRC64;

Query Match 10.3%; Score 518.5; DB 1; Length 872;
Best Local Similarity 29.2%; Pred. No. 3.5e-14;
Matches 276; Conservative 80; Mismatches 407; Indels 181; Gaps 54;

Oy 19 VYDEAGSGIDNDEKVFVTPDST--TOH-----NKVSTSKITTAAPINRPSLPPNS 69
Db 29 YGSAVSAAGSAGAYK-TLPGSHPGSKHVPYKPKNKIPT-PYI--SKSVAPAPKPKGY 84
Oy 70 DTSKETSIVNKEETVETKETTNTKOTSTGDEKETSIAKQSIKESASDIAPTSVL 129
Db 85 YPTKRYQPTYGSKTYPPYPIAKKLSYKAIKTYPAVAKTSYPSYK----HKIT 139
Oy 130 AKPTPKAETTKGPAITTPKKEPT--PTPKKPASTTPKEPTPTTIKSAPTTPKKEPAPTTPK 188
Db 140 YPTPYK-----PKITYP--PYKOKPSYPSYKPKTTPPYTK-----PKITYPPTYK 185
Oy 189 SAPT-TPKKAPAT--TKKAPATTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPK 243
Db 186 RKPSTTPYKPAATYPPYKPKITYP---PYKKRPSYK-----PYKKKTYPPYKPKI 236
Oy 244 PAPTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPK 296
Db 237 SYPSIYKPKASYVSYSKSKTYPPYKPKISYPPYKPKSYK-----PTYKPKATYPPTY 292
Oy 297 KKPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPK 351
Db 293 K-PKPSYPPYKPKITYPPT--YKPKPSYPPYKOKPSYPPYKSKSYPSYKSKTYPP 350
Oy 352 EPSPTTKKBPAPPTTKKBPAPPTTKKBPAPPTTKKBPAPPTTKKBPAPPTTKKBPAPPTTKK 401
Db 351 TYKPKITYPPTYKPKPSYPSYKPKKTYSPYKPKITYPPTYKPKPSYPSYKPKKTYPP 410
Oy 402 APVAPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPK 456
Db 411 T-YKPKISYPPYKPKASYVSYSKSKTYPPYKPKISYPPYKPKPSYPPYKPKKTYPP 469
Oy 457 PT-PTTPEEPAPTTPKKAANTPK-EPAPTTPKKEPAPTTP-----KBPAPTTP--PKETAP 507

```

Query Match	Best Local Similarity	Matches	Score	DB	Length
152	10.6%	29.4%	530.5	DB 1	1162
599	177: Conservative	100: Mismatches	212:	Indels	113: Gaps
211	KE-PA	PTTKEPA	PTTKEPA	PTTKEPA	PTTKEPA
652	ADSSAHG	PTSTPVDSSAHG	PTSTPVDSSAHG	PTSTPVDSSAHG	PTSTPVDSSAHG
258	KEPTT-PTTKEPA	PTTKEPA	PTTKEPA	PTTKEPA	PTTKEPA
712	VDSAHG	PTSTPVDSSAHG	PTSTPVDSSAHG	PTSTPVDSSAHG	PTSTPVDSSAHG
317	PTTKEPA	PTTKEPA	PTTKEPA	PTTKEPA	PTTKEPA
765	HGHP	PTSTPVDSSAHG	PTSTPVDSSAHG	PTSTPVDSSAHG	PTSTPVDSSAHG
373	TP-KRPA	PTTKEPA	PTTKEPA	PTTKEPA	PTTKEPA
818	TPVDSSAH	PTSTPVDSSAHG	PTSTPVDSSAHG	PTSTPVDSSAHG	PTSTPVDSSAHG
428	TP-EKLA	PTTKEPA	PTTKEPA	PTTKEPA	PTTKEPA
877	STPVDSSAH	PTSTPVDSSAHG	PTSTPVDSSAHG	PTSTPVDSSAHG	PTSTPVDSSAHG
478	PKERAP	PTTKEPA	PTTKEPA	PTTKEPA	PTTKEPA
936	PSRPA	PTTKEPA	PTTKEPA	PTTKEPA	PTTKEPA
538	TTKEP	PTTKEPA	PTTKEPA	PTTKEPA	PTTKEPA
984	PTSTPVDSSAH	PTSTPVDSSAHG	PTSTPVDSSAHG	PTSTPVDSSAHG	PTSTPVDSSAHG
597	KPA	PTTKEPA	PTTKEPA	PTTKEPA	PTTKEPA
1034	TP----	ADSSAHG	PTSTPVDSSAHG	PTSTPVDSSAHG	PTSTPVDSSAHG
653	PTTSEV	PTTKEPA	PTTKEPA	PTTKEPA	PTTKEPA
1085	DSSAH	PTSTPVDSSAHG	PTSTPVDSSAHG	PTSTPVDSSAHG	PTSTPVDSSAHG
712	MT	713	MT	713	MT
1126	GT	1127	GT	1127	GT

```

DE CALPHOTIN.
CN CN OR CAP.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;
OC Ephydraidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CANTON-S;
RX MEDLINE=93165729; PubMed=8094559;
RA Mattin J.H., Benzer S., Rudnicka M., Miller C.A.;
RT "Calphotin: a Drosophila photoreceptor cell calcium-binding protein.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:1551-1555(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CANTON-S;
RX MEDLINE=93165730; PubMed=8434015;
RA Ballinger D.G., Xue N., Harshman K.D.;
RT "A Drosophila photoreceptor cell-specific protein, calphotin, binds
RT calcium and contains a leucine zipper."
RL Proc. Natl. Acad. Sci. U.S.A. 90:1536-1540(1993).
CC -1- FUNCTION: MIGHT FUNCTION AS A CALCIUM-SEQUESTERING "SPONGE" TO
CC REGULATE THE AMOUNT OF FREE CYTOPLASMIC CALCIUM. IT BINDS 0.3 MOL
CC OF CA+2 PER MOL OF PROTEIN.
CC -1- SUBUNIT: HOMODIMER (PROBABLE).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC; HYPODENSE COMPARTMENT.
CC -1- TISSUE SPECIFICITY: SOMA AND AXONS OF PHOTORECEPTOR CELLS OF
CC COMPOUND EYES AND OCELLI.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED EARLY IN PHOTORECEPTOR CELL
CC DEVELOPMENT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by, and for commercial/
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L02111; AAA28405.1; -.
CC DR EMBL: L05080; AAA28420.1; -.
CC DR PIR: A47282; A47282.
CC Flybase: FBgn0010218; Cpn.
KW Calcium-binding.
FT CONFILCT 36 36 A -> AVAPAVVA (IN REF. 2).
FT FT 43 43 I -> T (IN REF. 2).
FT FT 64 64 I -> V (IN REF. 2).
FT FT 76 76 P -> A (IN REF. 2).
FT FT 100 100 P -> P (IN REF. 2).
FT FT 126 127 VO -> AP (IN REF. 2).
FT FT 154 154 I -> V (IN REF. 2).
FT FT 160 160 S -> T (IN REF. 2).
FT FT 534 534 A -> E (IN REF. 2).
FT FT 699 699 I -> T (IN REF. 2).
FT FT 703 703 V -> L (IN REF. 2).
FT FT 721 721 D -> E (IN REF. 2).
FT CONFILCT 721 721
SQ SEQUENCE 865 AA; 84781 MW; 2110417E0BE07CFE CRC64;

Query Match 10.6%; Score 530; DB 1; Length 865;
Best Local Similarity 26.2%; Pred. No. 1.2e-14;
Matches 233; Conservative 73; Mismatches 433; Indels 152; Gaps 411

OY 123 APTSKVLAKD-TPK---AETTTKGPAITTPKEPTPTTPKPEASATTTKEPEPTTIKASPTT 178
DB 8 SPSPASPAVAEVTSAVAAPQOVASPAAVP-----APAPAIIVTVPVAPPPTLASVOAPT 61
OY 179 PKBPAPTTTSAPPTP-KEBPATTTKEBPATTTKEBPATTTKEBPAPT-----TTKSAPPTP 233
DB 62 VTIPAPPAIAASVTIPVASVAPPVAAFP-----PAASVSTIPVAVAQIPVASVAPVAP 116
OY 234 KEBPATPKKPPAP-----TTKBPAPTTTKBPTP-----TTKBPAPTTKEBPATTT 279

```

DR Pfam: PF00038; filament; 1.
 DR PROSITE: PS00226; IF; 1.
 KW Intermediate filament; Coiled coil; Neutrone; Phosphorylation;
 Repeat:
 FT DOMAIN 1 97 HEAD.
 FT DOMAIN 98 408 ROD.
 FT DOMAIN 409 1087 TAIL.
 FT DOMAIN 436 517 GLU-RICH (ACIDIC).
 FT DOMAIN 519 886 50 X 6 AA TANDEM REPEATS OF K-S-P-A-E-A.
 FT DOMAIN 887 1087 GLU/LYS-RICH.
 FT DOMAIN 98 129 COIL 1A.
 FT DOMAIN 130 141 LINKER 1.
 FT DOMAIN 142 239 COIL 1B.
 FT DOMAIN 240 261 LINKER 12.
 FT DOMAIN 262 283 COIL 2A.
 FT DOMAIN 284 287 LINKER 2.
 FT DOMAIN 288 408 COIL 2B.
 FT CONFLICT 133 133 K -> QA (IN REF. 2 AND 3).
 FT CONFLICT 139 199 A -> AR (IN REF. 2 AND 3).
 FT CONFLICT 199 281 S -> T (IN REF. 2 AND 3).
 FT CONFLICT 281 492 L -> G (IN REF. 2 AND 3).
 FT CONFLICT 492 551 P -> PREAKSP (IN REF. 3).
 FT CONFLICT 551 712 MISSING (IN REF. 3).
 FT CONFLICT 689 712 G -> A (IN REF. 3).
 FT CONFLICT 714 714 T -> N (IN REF. 2 AND 3).
 FT CONFLICT 814 814
 FT CONFLICT 843 843
 SQ SEQUENCE 1087 AA: 116612 MW: 57BAC76A38EDICB9 CRC64;

Query Match 10.6%; Score 533; DB 1; Length 1087;
 Best Local Similarity 27.5%; Pred. No. 1,1e-14;
 Matches 216; Conservative 81; Mismatches 330; Indels 158; Gaps 38;

QY 21 DEAGSGLDNGDFKTY-----TPDTSTQHNKVTSPKTTAKPIINPRSLPNDTSKET 75
 DB 404 ECHIGGSPFSLTEGLPKIPISIT--HIVKSEMIKVE-----KSEKET 449
 QY 76 SLVWKETVYTKETTTNNKOTSTGDEKTTSAKETOSIEKTSKDLAPTSKVLAKPTPK 135
 DB 450 VIVEGQEEELRVTEGVVEEDKEAOGEGEEAEDEEELAAATSPAEAAPEKE 509
 QY 136 AETTKGPALTPKREPTTPKPEASPTTPKPTTIKSNATTPKKEAPTTKSAPTTPK 195
 DB 510 TKSANKEEAKSPGAKEAKSPA-----EAKSGEAKS--PGEAKEAKSPAEPKSPA 564
 QY 196 EP-APTTKEAPTTKPEAPTTKPE-----APTTKSAPTTKPEPA-----PTPKKPA- 245
 DB 565 EPKSPAEAKSPA--EPKSPA--TVKSPGEAKSPSEAKS--PAEAKSPAEEAKSPAEPK 619
 QY 246 ---PTTKEPA-----PTTKEP-----TPTTKEAPTTKPEAPTTKPEPA-----PTAPKK 290
 DB 620 AKSPAFAKSPAEAKSPATVNSPGEAKEAKSPA-----EAKSPAFAKSPAEAKSPA 676
 QY 291 P---AETTKPEAPTTKPEAPTTKPEAPTTKPEPA-----PTTTSAPTTKPEAPT 342
 DB 677 PGEAKEAPKPEAKSPA--KSPA-----EVKSPAFAKSPAFAKSPGEAKS--PAAYKSPA 729
 QY 343 TKSAPTTKPE--PSTTTTKEPA-----PTTKEAPTTKPEAPTTKPEAPTTKPEAPT 397
 DB 730 SPAAYKSPGAKEAKSPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSPA 769
 QY 398 TKRPAAPTAPEAPTTKPEAPTTKPKLITTPTEKLAPTTPEKPAPTTPELAPTTPEEP 457
 DB 790 AKSPVAKEDIPPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSPA 848
 QY 458 TPTTPEE--PAPTTKPAAPTTKPEAPTT-----KPEAPTT-----KEPAPTTPEK 505
 DB 849 RP--PEGVKSPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSPA 904
 QY 506 APPTPKGAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPE 563
 DB 905 TLTPTTPEAKSKDEAPKPEAPKVEEKEETPTKEPKDSTAFAKPEAKKEAKKAAVASEE 964

QY 564 KEPAPTTKEPAPTTTPKGAPTTKEAPAPTTKPEAPKPAELAPTTTKGPTSTSDKPAFTT 623
 DB 965 ETPAKLVKVEEA--KPKETETTTKEADTKAKES-----KPTKEK----- 1006
 QY 624 PRETAPTTKEPAPTTTPKPAPTTETPTTSEVSTPTTKEPTTIKSPDESTEPELSA 683
 DB 1007 -----KKEMPAPAK-----KDTKEKTTESKPEEK----- 1034
 QY 684 EPTPALENSPEKPEGVPTTTPATKPEMTTAKDTERDLRTTPTTAAPKMTKETA 743
 DB 1035 ---PK--MEAKVKEDDKSLSKEP--SKPTKEAKSSSIDQNESQPE-----KTTEDKA 1082
 QY 744 TTTEK 748
 DB 1083 TKGEK 1087

RESULT 9
 ID TCNA_TRYCR STANDARD; PRT: 1162 AA.
 AC P23253;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE STRIDASE (EC 3.2.1.18) (NEURAMINIDASE) (NA) (MAJOR SURFACE ANTIGEN).
 GN TCNA.
 OS Trypanosoma cruzi.
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
 OX NCBI_Taxid=5693;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-STIVIO X-10/4;
 RX MEDLINE=91277609; PubMed=1711561;
 RA Pereira M.E.A., Mejia J.S., Ortega-Barria E., Matzilevich D.,
 RA Prioli R.P.;
 RT "The Trypanosoma cruzi neuraminidase contains sequences similar to
 RT bacterial neuraminidases, WMD repeats of the low density lipoprotein
 RT receptor, and type III modules of fibronectin.";
 RL J. Exp. Med. 174:179-191(1991).
 RN [2]
 RP SUBCELLULAR LOCATION.
 RX MEDLINE=91376547; PubMed=1896773;
 RA Prioli R.P., Mejia J.S., Ajl T., Aikawa M., Pereira M.E.A.;
 RT "Trypanosoma cruzi: Localization of neuraminidase on the surface of
 RT trypanomastigotes.";
 RL Trop. Med. Parasitol. 42:146-150(1991).
 CC - FUNCTION: DEVELOPMENTALLY REGULATED NEURAMINIDASE IMPLICATED IN
 CC PARASITE INVASION OF CELLS.
 CC - CATALYTIC ACTIVITY: HYDROLYSIS OF 2,3-, 2,6- AND 2,8-GLYCOSIDIC
 CC LINKAGES JOINING TERMINAL NON-REDUCING N- OR O-ACETYLRAMINYL
 CC RESIDUES TO GALACTOSE, N-ACETYLHEXOSAMINE, OR N- OR O-ACETYL
 CC NEURAMINYL RESIDUES IN OLIGOSACCHARIDES, GLYCOPROTEINS,
 CC GLYCOLIPIDS OR COLOMINEIC ACID.
 CC - SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
 CC (POSSIBLE).
 CC - DEVELOPMENTAL STAGE: MAXIMAL ACTIVITY IN TRYPANOMASTIGOTES, MINIMUM
 CC IN EPIMASTIGOTES AND NO DETECTABLE IN AMASTIGOTES.
 CC - MISCELLANEOUS: THE VARIABLE LENGTHS OF THE LONG TANDEM REPEAT
 CC DOMAIN COULD ACCOUNT IN PART FOR THE POLYMORPHISM OF THE TCNA
 CC PROTEIN.
 CC - SIMILARITY: BELONGS TO FAMILY 33 OF GLYCOSYL HYDROLASES.
 CC - SIMILARITY: CONTAINS 3 BNR REPEATS.
 CC -
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch)
 CC EMBL: M61732; AAB30255.1; -
 DR PIR: JH0557; JH0557.

CC -1- PTM: EXTENSINS CONTAIN A CHARACTERISTIC REPEAT OF THE PENTADEPTIDE
 CC SER-PRO(4). THE PROLINE RESIDUE IS HYDROXYLATED AND THEN
 CC GLYOXYLATED.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X13885; CAA32090.1; -
 CC DR PIR: S06733; S06733.
 CC KW Repeat: Cell wall; Glycoprotein; signal; structural protein;
 KM Hydroxylation.
 FT SIGNAL 1 ?
 FT CHAIN ? 620 EXTENSIN.
 FT REPEAT 70 73 H-A-P-P.
 FT REPEAT 148 151 H-A-P-P.
 FT DOMAIN 229 242 2 X 7 AA TANDEM REPEATS OF T-H-R-H-A-P-P.
 FT REPEAT 229 235 1.
 FT REPEAT 236 242 2.
 FT DOMAIN 205 620 CONTAINS THE SER-PRO(4) REPEATS.
 FT DOMAIN 499 600 3 X APPROXIMATE TANDEM REPEATS.
 SQ SEQUENCE 620 AA: 65406 MW: 641DD2270AB28524 CRC64;

Query Match 11.0%; Score 551; DB 1; Length 620;
 Best Local Similarity 27.2%; Pred. No. 1,5e-15;
 Matches 170; Conservative 64; Mismatches 304; Indels 88; Gaps 24;

QY 136 AETTK-GPALTP--KEPTT-----PEKASTPKE--PPPTIKSAPTTKEAP 184
 DB 24 AEATTOYGGLPRPVTSQPPSSIGLSPPSAPPTTPSRGHNVPSP--RNAHPNHAAPP 80
 QY 185 TTTSKAPTTER-----APTTEKAPAT--TPKEAPPTTKEAPATTTKSAPTTKE 235
 DB 81 SHGLPRSVGPRPHRGNHLPSSKGFNPPSPVVISHPNPPSYGAPRPSHGHLPSHCQR 140
 QY 236 PAPTTPKPPAP--TPKEAPATTTKEAPATTTKEAPATTTKEAPATTTKEAPATTTK 290
 DB 141 P-----PSPSHGNAPRRSGGHTPRGONHRRSPSRSHGNAPRTYQAQPPRTIYS 193
 QY 291 PAPTTPKPPAPTTKEAPATTTKESPTTK--EPATTTKSAPTTKEAPATTTK-- 345
 DB 194 PSPGV--QPPPT--YSPRPTNVOPTSPSRGHOPOPTTHNAPRTNHOBSPL 249
 QY 346 ---APTTPKESPTTKEAPATTTKEAPATTTKPPAPTTKEAPATTT--TPKEAPATTTK 399
 DB 250 RHLPPSPRROPTTSPPPAYASQPSPTYSPPPTYSPPPPYSPPPPAYSSP 309
 QY 400 KPAAT--APKEAPATTTKETAPTTPKL-TPTT-----PEKAPTTKEAPATTTPEL 449
 DB 310 PPTLPPTSPPPAYSPPTTSPPTTLPPLSPSPTYSPPPPVYSPPPPSYSPPTTTL 369
 QY 450 APTTPKEPTTPEAPATTTKAAAP--NTPKEAPATTTKEAPATTTKEAPATTTKETA 506
 DB 370 PPPSPSPSPSPSPPTTQSPSPPPAYSPPLAPPTYSPPPT--YSPPPPTIYAQPP 427
 QY 507 PTTKGAATTLKEAPATTTKPPAPKELAPTTTKEPTTSTSDKRAPTTPKGTAPTTKEP 566
 DB 428 P-----LPPTYSPPPPAYSPPTTSPPTTSPPPAYASQPPPPPTYSPPPPAYSP 481
 QY 567 APTTPKEAPATTTKGTAPTTTKEAPATTTKPPAPKELAPTTTGTSTSDKRAPTTKE 626
 DB 482 PPPSPYSPPPVQVPLPPTSPSPRRRIHLPPRHOP--RPPTTPTGQPPSPPTSP 538
 QY 627 TAPTTKPPAP--TPKAPPTTPEPTTSEVSTPTTKEPTTIHNSPDSSTELSAE 684
 DB 539 PPPROIHSPPHNQPIPTTPTGQPPSPPTTSPAPPPROIHSPPHNQPRPTTPTTQCP 598
 QY 685 PTPKALENSPKRPGVPTTKTPATKP 710

DB 599 PSP-----PTTYSPPSPPP 612

RESULT 8
 ID NPH_MOUSE STANDARD; PRT; 1087 AA.
 AC P19246; 061959;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE NEUROFILAMENT TRIPLET H PROTEIN (200 KDA NEUROFILAMENT PROTEIN)
 DE (NEUROFILAMENT HEAVY POLYPEPTIDE) (NF-H).
 GN NPH OR NPH.
 OS Mus musculus (mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89121513; PubMed=3220257;
 RA Jullien J.-P., Cote F., Beaudet L., Sidky M., Flavell D., Grosveld F.,
 RA Mushynski W.;
 RT "sequence and structure of the mouse gene coding for the largest
 RT neurofilament subunit.";
 RL Gene 68:307-314(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89089138; PubMed=3145094;
 RA Sheldman P.S., Gaden M.J., Lees J.F., Lazzarini R.A.;
 RT "The structure of the largest murine neurofilament protein (NF-H) as
 RT revealed by cDNA and genomic sequences.";
 RL Brain Res. 464:217-231(1988).
 RN [3]
 RP SEQUENCE FROM N.A.
 RP STRAIN=SWISS WEBSTER; TISSUE=Brain;
 RA Gaden M.J.;
 RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
 CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
 CC NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT
 CC OBSERVED BY THE TWO SMALLER NF PROTEINS.
 CC -1- PHOSPHORYLATION: THERE ARE A NUMBER OF REPEATS OF THE TRIPLET K-S-P. NPH IS
 CC THOUGHT THAT PHOSPHORYLATION OF NPH RESULTS IN THE FORMATION OF
 CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
 CC OF AXONAL CALIBER.
 CC -1- PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
 CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H). THE
 CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
 CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
 CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
 CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN IN POSITIONS 534
 CC TO 716 AND IS SHORTER DUE TO FRAMESHIFTS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M24496; AAA39813.1; -
 CC EMBL: M24496; AAA39813.1; JOINED.
 CC EMBL: M24494; AAA39813.1; JOINED.
 CC EMBL: M24495; AAA39813.1; JOINED.
 CC EMBL: M35131; AAA39809.1; ALT_FRAME.
 CC EMBL: M35102; CAA83229.1; -
 CC PIR: J03038; QFMNH.
 CC PIR: A43778; A43778.
 CC MGD: MGI:97309; NfH.
 CC InterPro: IPR001664; IF.

RT protein.";
 RL Biotechnol. Prog. 6:171-177(1990).
 RN PARTIAL SEQUENCE, AND POST-TRANSLATIONAL MODIFICATIONS.
 RP MEDLINE=8135732; PubMed=6298211;
 RX Waite J.H.;
 RA "Evidence for a repeating 3,4-dihydroxyphenylalanine- and
 RT hydroxyproline-containing decapeptide in the adhesive protein of the
 RL mussel, *Mytilus edulis* L."
 RL J. Biol. Chem. 258:2911-2915(1983).
 CC -I- FUNCTION: PROVIDES ADHESIVENESS TO THE MUSSEL'S FOOT. MUSSELS
 CC PRODUCE ONE OF THE STRONGEST WATER INSOLUBLE GLUE. THE MUSSEL'S
 CC ADHESIVE IS A BUNDLE OF THREADS, CALLED A BYSSUS, FORMED BY A
 CC FIBROUS COLLAGENOUS CORE COATED WITH ADHESIVE PROTEINS.
 CC -I- SUBCELLULAR LOCATION: SECRETED.
 CC -I- TISSUE SPECIFICITY: PRODUCED BY THE BYSSAL GLAND.
 CC -I- DOMAIN: ALMOST EXCLUSIVELY COMPOSED OF REPEATS OF A DECAPEPTIDE.
 CC -I- PTM: THE DECAPEPTIDE A-K-P-S-Y-P-P-T-Y-K IS POST-TRANSLATIONALLY
 CC MODIFIED AS FOLLOWING: THE SIXTH AND SEVENTH RESIDUES ARE
 CC HYDROXYLATED AND THE PENULTIMATE IS A 3,4-DIHYDROXYPHENYLALANINE
 CC (DOPA) DERIVED FROM TYROSINE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X54422; CA38294.1;
 DR InterPro: IPR002964; Adhesive_plaq.
 DR InterPro: IPR002965; P-rich_extensn.
 DR PRINTS: PR01216; ADHESIVEI.
 DR PRINTS: PR01217; PRICHEXTENS.
 DR Repeat: Hydroxylation.
 FT NON_TER 1
 FT DOMAIN 67 870
 FT TANDEN REPEATS OF Y-K-[PA]-K-[LP]-[ST]-Y-
 FT P-P-[ST]
 FT SEQUENCE 875 AA; 100412 MM; 6EAB5312748CAACE CRC64;
 SQ
 Query Match 11.1%; Score 556.5; DB 1; Length 875;
 Best Local Similarity 28.7%; Pred. No. 1.2e-15;
 Matches 272; Conservative 106; Mismatches 375; Indels 195; Gaps 56;
 QY 42 TQHNKYSTSPKITAKPINDP-----RP--SLPNSDSKETSILTVNKEETVET 87
 DB 1 TKEHPYKPKKTSYSAKYKPYQPLKKVDRPTKSYPTYG-SKTNYLPLAKLSSYK 59
 QY 88 KETTTNKTSDG--KEKTT---SAKETOSIEKTSKDLAPTSKYLAKEPTPAKETTTCG 142
 DB 60 PIKTYNAKTNPPVYKPKMTYPPYKPKPSYPTKSKPTKYKPTTYPTAKKPSY-- 117
 QY 143 PALTTPKEPTTPKE-----PASTPKEPTPTTIKSAPTTP-----KEPATTTTKSAPTT 193
 DB 118 PSSYKPKKTPPYKREKLTPPYKPKPSYPTKPKPSYPTKPKPSYPTKPKPSY 177
 QY 194 PKEPAPTTPKEA--PTTPKEPA-----PTTPKEPA--PTTTSKAPTTPK--EPAPT 240
 DB 178 P-----PTYKAKPSYPTPYKAKPSYPTPYKAKPSYPTPYKAKPSYPTPYKAKPSY 233
 QY 241 PKKPA--PTTPKEPA--PTTPKEPT--PTTPKEP--APTTPKEAPTTPKEPA--PTAPK 289
 DB 234 KAPSPYPTPYKAKPSYPTPYKAKPSYPTPYKAKPSYPTPYKAKPSYPTPYKAKPSY 293
 QY 290 KPA--PTTPKEPA--PTTPKEPA--PTTPKEPS--PTTPKEPA--PTTTSKAPTTPKEPA-- 339
 DB 294 KPSYPTPYKAKPSYPTPYKAKPSYPTPYKAKPSYPTPYKAKPSYPTPYKAKPSY 353
 QY 340 PTTTSKAPTTPKEPSPTTPKEAPTTPKEAPTTPKEAPTTPKEPA--PTTPKEPA--PTTPKEPA 395
 DB 354 PSYKAKPSYPTPYKAKPSYPTPYKAKPSYPTPYKAKPSYPTPYKAKPSYPTPYKAKPSY 409

QY 396 TTTKKPAPTAKPEAPTTPKEAPTTPKLTPTTPEKLAAPTTPKEKAPTTPBEIAPTTPPE 455
 DB 410 PPTTKAKPTTYKAP-----TYPST-YKAPSY-----PSYKAKPSYPTPYKAKPSYKA 457
 QY 456 EPTTPPEEAPPTPYKKAAPNTPKPEAPTTPKEAPTTPKE-----PAPTTPKEAPTTP-- 509
 DB 458 KPT-----YPSYKA-----KPSYASYSKAKPSYPTPYKSKSSYSSSKPKKPYPTPYK 506
 QY 510 PKGTAPTLKEAPPTTPKPAKELAPTTTKEPTSTSDKAPAPTTPKGTAPTTP--PKE 565
 DB 507 PKLTYPKPTPK--PKPSYPSYKKTTPPYPTPK-----PKISTYPTPYKAKSY 551
 QY 566 PAPTTPKEAPTTPKGT--APTTPKEPA--PTTPKEPA--PKELAPTTTPG-PTSTSD 617
 DB 552 PATTYKAKPSYPTPYKAKPSYPTPYKAKPSYPTPYKAKPSYPTPYKAKPSYPTPYK 611
 QY 618 KPAPTTPKEAPTTPKEPA--PTTPKEPA--PTT-----DETPEPTTSEVSTPTTKE----- 666
 DB 612 KAKPSYPT-----PTYKAKPSYPTPYKAKPSYPTPYKAKPSYPTPYKAKPSYPTPYK 667
 QY 667 PTTTKSPDESTPELSAETPKALENSK-----EGVPTT--KTPA-----ATKPEMTT 715
 DB 668 PPT-YKAKPSYPTPYKAKPSYPTPYKAKPSYPTPYKAKPSYPTPYKAKPSYPTPYK 726
 QY 716 AKDKTTERDLRTPETTTTAPKMTKETATTEKTESKITATTGVTSTTODTTPFKIT 775
 DB 727 YKAKPTPYKAKPSYPTPYKAKPSYPTPYKAKPSYPTPYKAKPSYPTPYKAKPSYPTPYK 777
 QY 776 TUKTTTLAPKVTYTKKTTITTEIMNKPETAKPKDRAINSKATTPKPKAPTAPK--P 832
 DB 778 TYKSKSIYSSYKPKKTPPT--YKPKLTTPPYK-----PKPSYPSYKPKKTIY 826
 QY 833 TSYKSKKTKMPRKPKKTPPT-----RKMTSTMPELNPSR 868
 DB 827 SYTKLKPSYPTPYKSKTSYPTPYKNTKISYSSYKAKTSIPPAKPTPNR 874
 RESULT 6
 ZAN_HUMAN STANDARD; PRT: 2700 AA.
 ID ZAN_HUMAN
 AC Q9Y493; O00218;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE ZONADHESIN (FRAGMENT).
 GN ZAN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE OF 1-2379 FROM N.A.
 RX MEDLINE=99018118; PubMed=9799793;
 RA Glockner G., Scherer S., Schatttevoy R., Boiligt A., Weber J.,
 RA Tsui L.C., Rosenthal A.;
 RT "Large-scale sequencing of two regions in human chromosome 7q22:
 RT analysis of 650 kb of genomic sequence around the EPO and CMT1 loci
 RT reveals 17 genes";
 RL Genome Res. 8:1060-1073(1998).
 RL [1]
 RN [1]
 RP SEQUENCE OF 2338-2700 FROM N.A.
 RC TISSUE=Testis;
 RX MEDLINE=97271566; PubMed=9126492;
 RA Cao Z., Harumi T., Garbers D.L.;
 RT "Chromosome localization of the mouse zonadhesin gene and the human
 RT zonadhesin gene (ZAN).";
 RL Genomics 41:119-122(1997).
 CC -I- FUNCTION: BINDS IN A SPECIES-SPECIFIC MANNER TO THE ZONA PELLUCIDA
 CC OF THE EGG. MAY BE INVOLVED IN GAMETE RECOGNITION AND/OR
 CC SIGNALING.
 CC -I- SUBUNIT: PROBABLY FORMS COVALENT OLIGOMERS.
 CC -I- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN, EXCLUSIVELY ON THE

DR	EMBL:	M61170;	AAB53150.1;	-	ALT_SEQ.
DR	EMBL:	X52229;	CAA36478.1;	-	ALT_SEQ.
DR	EMBL:	X52228;	CAA36477.1;	-	ALT_SEQ.
DR	EMBL:	M35093;	AAB59612.1;	-	ALT_SEQ.
DR	EMBL:	Z17344;	CAA78972.1;	-	-
DR	EMBL:	Z17345;	CAA78973.1;	-	-
DR	EMBL:	M31823;	AAAS35757.1;	-	ALT_INT.
DR	EMBL:	S81781;	AAD14376.1;	-	ALT_INT.
DR	EMBL:	M21868;	AAD14369.1;	-	ALT_INT.
DR	PIR:	A35175;	A35175.	-	ALT_SEQ.
DR	PIR:	B35175;	B35175.	-	-
DR	PIR:	S10218;	S10218.	-	-
DR	GlycoSuiteDB:	P15941;	-	-	-
DR	MIM:	158340;	-	-	-
DR	InterPro:	IIPRO00082;	SEA.	-	-
DR	Pfam:	PF01390;	SEA; 1.	-	-
DR	SMART:	SMO0200;	SEA; 1.	-	-
DR	PROSITE:	PS50024;	SEA; 1.	-	-
KV	Glycoprotein:	Signal;	Cytoskeleton; Actin-binding; Transmembrane; Repeat; Alternative splicing.	-	-
FT	SIGNAL	1	23	-	POTENTIAL.
FT	CHAIN	24	1255	-	MUCIN 1.
FT	DOMAIN	24	1162	-	EXTRACELLULAR (POTENTIAL).
FT	TANDMEM	1163	1186	-	POTENTIAL.
FT	DOMAIN	1187	1255	-	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	81	960	-	44 x 20 AA TANDEN REPEATS.
FT	DOMAIN	1034	1151	-	SEA.
FT	CARBOHYD	957	957	-	N-LINKED (GLCNAc . . .) (POTENTIAL).
FT	CARBOHYD	975	975	-	N-LINKED (GLCNAc . . .) (POTENTIAL).
FT	CARBOHYD	1029	1029	-	N-LINKED (GLCNAc . . .) (POTENTIAL).
FT	CARBOHYD	1055	1055	-	N-LINKED (GLCNAc . . .) (POTENTIAL).
FT	CARBOHYD	1133	1133	-	N-LINKED (GLCNAc . . .) (POTENTIAL).
FT	VASPPLIC	19	19	-	T -> TATTATPATP (IN ISOFORM B).
FT	VASPPLIC	20	22	-	MISSING (IN ISOFORM C).
FT	VASPPLIC	20	31	-	MISSING (IN ISOFORM D).
FT	VASPPLIC	126	905	-	MISSING (IN POLYMORPHIC EPITHELIAL ISOFORM).
FT	VASPPLIC	1077	1087	-	EIOIYKCGFL -> VSIGISFPMPLP (IN SECRETED ISOFORM).
FT	VASPPLIC	1088	1255	-	MISSING (IN SECRETED ISOFORM).
FT	CONFLICT	2	2	-	T -> A (IN REF. 11).
FT	CONFLICT	134	134	-	P -> Q (IN REF. 9).
FT	CONFLICT	154	154	-	P -> Q (IN REF. 9).
FT	CONFLICT	1021	1021	-	S -> T (IN REF. 3).
FT	CONFLICT	1251	1251	-	A -> T (IN REF. 3).
SO	SEQUENCE	1255 AA;	122072 MW;	5E2BDFC4DE57DA82 CRC64;	-

Query Match	13.0%;	Score 651;	DB 1;	Length 1255;
Best Local Similarity	27.1%;	Pred. No. 3.4e-19;		
Matches 290;	Conservative 79;	Mismatches 437;	Indels 266;	Gaps 52

133 1SLVLAIVEKRETTTNNKQTSIDGKEKTSKETSKEIETSIKTSAKD-LAPTSKVLAKPP
 Db 16 TVLTV-----VTGSGHASSPGEKEKETSATQSSVSPSTREKNAVENTISVLSHIS 65
 Qy 134 P-KAETTTKGP--ALTTPKEP-----TPTTPKEPASTP-----KE 166
 Db 66 PGSGSSITGGQDVTLAPATERPAGSGSAATGQDVTSVPTPRALGISTPRAHDVTSAPDNK 125
 Qy 167 PTPTT-----IKSAPTTPKPEAPTTT-----KSAPTTPKPEAPTTTKEP-----A 206
 Db 126 PARGSTAPRAHGVTSAPDT--RPARGSTAPRAHGVTSAPDT--RPARGSTAPRAHGVTS 181
 Qy 207 PTPPKPEAPTTTKEBAPTTTKSAPTTPKPEAPTTT-----KKAPTTPKPE----- 252
 Db 182 PDL--RPARGSTAPRAHGVTSAPDTPRAPGSTAPRAHGVTSAPDTPRAPAGSTAPRAHGV 238
 Qy 253 --APTPKPEPTPTT-----KEBAPTTKEP-----APTPKPEAPTPK----- 289
 Db 239 TSAPDTPRAPAGSTAPRAHGVTSAPDTPRAPAGSTAPRAHGVTSAPDTPRAPAGSTAPRAHGV 298

QY	290	KPATTTPKEAPPTPK-----EPAPTTTKEESPPTTKEBAPTT---TKSAPTTTKEBAPTT	34
Db	299	TSAPDTPRAGSGTAPRPAHGVTSAPDTPRAGSGTAPRPAHGVTSAPDTPRAGSGTAPRPAHGV	35
QY	343	TKSAPPTTKEESPPTTKEP-----APTTTKEBAPPTPK-----KPAPTTKEBAPPTPK-	39
Db	359	T-SAPDT--RPAAGSTAPRPAHGVTSAPDTPRAGSGTAPRPAHGVTSAPDTPRAGSGTAPRA	41
QY	392	---EPAPTTTKKAPRPAKKEBAPTTTKEBAPTTTPKKLTPTTEBKAAPTTPEKAPTTPEE	44
Db	416	HGVTSAPDTPRAGSGTAPRPAHGVTSAPDTPRAGSGTAPRPAHGVTSAPDTPRAGSGTAP--	47
QY	449	LAPTEEBEPPTTPEBAPPTTKEA-----AAPTTTKEBAPPTTPK-----BPAPTTKEBAPT	50
Db	474	--PAHGVTSAPDTPRAGSGTAPRPAHGVTSAPDTPRAGSGTAPRPAHGVTSAPDTPRAGSGT	53
QY	501	TPKETAPPTPKGAPPTTKEBAPPTP-----KKPAKELAPTTTKEBPTTSDKP	55
Db	532	AP----PAHGVTSAPDTPRAGSGTAPRPAHGVTSAPDTPRAGSGTAP--PAHGVTSAPDTPR	58
QY	551	AP--TPPKGAPPTTKEBAPPTTKEBAPPTTKEBPTTKEBPTTKEBPTTKEBPTTKEBPTTKEB	59
Db	587	APGSTAPRPAHGVTS-----APDTPRAGSGTAPRPAHGVTSAPDTPRAGSGTAPRPAHGVTSAP	64
QY	596	--KKPAKELAPTTTKEBPTTSDKPAP--TTPK-----ETAPTTKEBAPPTP-----	64
Db	643	DTPRPAAGSTAP--PAHGVTSAPDTPRAGSGTAPRPAHGVTSAPDTPRAGSGTAPRPAHGVTSAP	70
QY	641	--KKPAPTTPEPTTPEPTTPEPTTPEPTTPEPTTPEPTTPEPTTPEPTTPEPTTPEPTTPEPTT	69
Db	702	PTDTPRAGSG--TAPRPAHGVTSAPDTPRAGSGTAPRPAHGVTSAPDTPRAGSGTAPRPAHGV	75
QY	692	NSPKRPGVPTTTPRA-----TKPMTTAAADK--TTERDLRTTPEPTT-----T	73
Db	760	SAPDTPRAGSGTAPRPAHGVTSAPDTPRAGSGTAPRPAHGVTSAPDTPRAGSGTAPRPAHGV	81
QY	734	AAPKTKETATTTTETKTESKITATTTTQVTSITTDTP--FKITLTKTTTAPRVTTTCK	79
Db	820	SAP-----DTPRAGSGTAPRPAHGVTSAPDTPRAGSGTAPRPAHGVTSAPDTPRAGSGTAPRA	87
QY	792	TITTT--EIMMKPEETAAPKDRATNSKATTPPKPOKPTTAPKPKPTTCKPTTAPVRPKPT	84
Db	876	HGVTSAPDTPRAGSGTAPRPAHGVTSAPDTPRAGSGTAPRPAHGVTSAPDTPRAGSGTAPRPAH	93
QY	850	TPTPKMTSTMBELNP-----TSKIAEMALO--TTPTPNOTPNSK--	88
Db	931	TAPRPAHGVTSAPRPNKRALGSTARPVHNVTSASGSASGSASASTLVINGNSARATTTTPASKST	99
QY	888	-----LVEVNPKESEDAGAGEETPTMILRLRPVFMPEVTPPMDY	92
Db	991	PSIPSHSDPTTTLASHSTKTADASTTHNSVPLTSSNHTSPOLSTGVSF	104

[illegible]

QY 871 EAMLTTPRNPQNSKLVENPKSEDAGAGETPPHMLLRPHVMPREVTPOMDXLP 927
DB 1425 -----KPTSTPAPTEIE-EPTPSDVGATIGEHRAVIRGCI-----PDGSEFP 1465

RESULT 3
AMYL_YEAST
ID AMYL_YEAST STANDARD; PRT: 1367 AA.
AC P08640; P08068;
DT 01-APR-1988 (rel. 08, Created)
DT 01-FEB-1995 (rel. 31, Last sequence update)
DT 20-AUG-2001 (rel. 40, Last annotation update)
DE GLUCOMYLASE S1/S2 PRECURSOR (EC 3.2.1.3) (GLUCAN 1,4-ALPHA-
GN STAI OR STAI2 OR MAL5 OR YIR019C.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RA Churchill B.G., Badcock K., Bankier A.T., Bowman D.,
RA Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA Goulet S., Hamlyn N., Hornell T.S., Hunt S., Jagels K., Jones M.,
RA Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
RA Rajendram M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA Walsh S.V., Whitehead S.;
RL Submitted (DEC-1994) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-242 AND 762-1331 FROM N.A.
RX MEDLINE=87194600; PubMed=3106330.
RT "Gene fusion is a possible mechanism underlying the evolution of
RT STAI...";
RL J. Bacteriol. 169:2142-2149(1987).
RN [3]
RP SEQUENCE OF 1-31 FROM N.A.
RX STRAIN-SPX101-1C;
RX MEDLINE=89031230; PubMed=3141213;
RA Pardo J.M., Ianez E., Zalacain M., Clarios M.G., Jimenez A.;
RT "Similar short elements in the 5' regions of the STAI2 and SCA genes
RT from Saccharomyces cerevisiae.";
RL FEBS Lett. 239:179-184(1988).
CC -1 CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL 1,4-LINKED ALPHA-D-
CC GLUCOSE RESIDUES SUCCESSIVELY FROM NON-REDUCING ENDS OF THE CHAINS
CC WITH RELEASE OF BETA-D-GLUCOSE.
CC -1 SIMILARITY: TO S.POMBE SPCC215.13.
CC -1 SIMILARITY: SOME, TO S.POMBE SPCC285.13C.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: Z38061; CA86176.1; -
DR EMBL: M16164; AAA35014.1; -
DR EMBL: M16165; AAA35015.1; -
DR EMBL: X13857; CAA32069.1; -
DR PIR: B26877; B26877. -
DR PIR: A26877; A26877. -
DR PIR: S48478; S48478. -
DR SGD: S0001458; MUC1.
KM Hydrolyase; Glycosidase; Polysaccharide degradation; Glycoprotein;
KM Signal; Multigene family.
FT SIGNAL 1 21
FT CHAIN 22 1367 POTENTIAL.
FT DOMAIN 210 1367 SER/THR-RICH.
FT CARBOHYD 817 817 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 874 874 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1367 AA; 136110 MW; 91C00E2DBD61A9D CRC64;

Query Match 15.28; Score 762.5; DB 1; Length 1367;
Best Local Similarity 27.88; Pred. No. 17e-23;
Matches 280; Conservative 115; Mismatches 441; Indels 171; Gaps 42;

QY 35 TTPDTSTTQHNKVVSTPKITAKPINP-----RPSLPNSDTSKETSILVKEETVE 86
Db STSESTTSSSTSSSTSSSTTAIPATITTCIEKEPPTTCTIEKPPPHHDTPPC 295
QY 87 TKETITTKOSTOCKETTSKAKETOSTEKTANDLAPTSKVLAKPPKAEPTT--TRGPA 144
Db TTKKTTTTSK-TCT--KTTTPVPPTPS--STTESSSAPV-----PPSSSTTESSAPV 344
QY 145 LTPKE-----PPTTPKEPASTTPKEPPTTTSKAP-----TPKKEPAPTTKSAFTTP 194
Db TSTTESSAPVPTPSSSTTESSAPVTSSTTESSAPVTSSTTESSAPVPTPSSSTTE 404
QY 195 KEPA-----TTKEPAP--TTKEPAP--TTKEPAP--TTKSAFTTPKEPAP 238
Db SSSAPVTSSTTESSAPVTSSTTESSAPVTSSTTESSAPVTSSTTESSAPVPTPSS 464
QY 239 TTPKKPAPTPKEPAPTTKEPPTT--TTKEPAPTTKE-----PAPT-----TPKEPA 284
Db TTESSAPVTSSTTESSAPVPTPSSSTTESSAPVTSSTTESSAPVTSSTTESS 524
QY 285 PTAKKAPAPTPKEPAPTTKE-----PAPTTKEPPTTTPKEPAPTTKSAFTTP 336
Db APAPTSSSTTESSAPVTSSTTESSAPVPTPSSSTTESSSTTPVTSSTTESSAPVPT 583
QY 337 EPAPTTPKSAFTTPKEPPTTTPKEPAPTTTPKEPAPTTTPKAPAPTT-----PREPAPTP 390
Db -PSSSTTESSAPVPTPSSSTTESSAPAPTPSSSTTESSAPAPTPSSSTTESSAPVTS 640
QY 391 KEPAPTTTPKAPAPTPKEPAPTTTPKEPAPTTTPKAPAPTTTPKEPAPTTTP 446
Db -PSSSTTESSAPVPTPSSSTTESSAP--VPPTSSSTTESSAPVTSSTTESSAPVTS 697
QY 447 ELAPTTPEEPTP--TTPEEAPPT--TPKAADNTPKEPAP--TTPEEAPPT--TT 493
Db STTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVTS 757
QY 494 -----PKEPAPTTKEPAPTT--TPKGTAPTTKEPAPTTTPKAPAPTTTPKPTST 545
Db ESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVPTPSSSN 817
QY 546 TSDKAPAPTPPKGAPTPTPKEPAPTTTPKEPAPTTTPKGTAPTTTPKEPAPTTTP 605
Db ITSSAPSTPSSSTTESSAPVPT--PSSSTTESSAPVTSSTTESSAPVPTPSSSN 874
QY 606 TTKKGP-----TSTDKAPATTPKETAFTTPKEPAPTTTPKAPAPTTPT--PPTTSEV 658
Db ITSSAPSTPSSSTTESSSTGT--VPSSSKYPSQETSVSTTEETVTPRTTTSV 931
QY 659 SPTTTPKEPTTIHKSPEDESTPELSAEPKALENSPKREGVTTTPATKREMT----- 714
Db TTPSTTTITTVCSGTGNSAGETGSCSKVYTT-----VTTTTTSTTSTTTT 986
QY 715 ---TAKDKTTEDLRTPEPT--TTAAPKMTKETATTEKTESKITATTOVST--TTOD 768
Db VCSGTGNSAGETGSCSKPTITTVPCSPSETASESTSPPTPVTTVSTVTTVEY 1046
QY 769 TTPFK-----ITTLKTTTLA--PKVTT-----TKKTTTTTEL--MKPEET- 805
Db STSTPAGGEITTTVTKINPTIYLLTIAPTSVITVTNFTPTTITTVCSGTGNSAGET 1106
QY 806 --AKKDRATNSKATTPKAPQKTAAPKPKPTSKKPKTPVRKPKPTTPTPPKMTSTMBEL 863
Db SCGSKRTVYTVPCSTGGEYTTAATLVAVTTVTTSTSGTNSAGCTTGGYTTKS 1106
QY 864 NPTSRIAEAMLTTPRNPQ--TPNSKLVENPKSED--ACGAGET 905

	Cell wall, S-layer; Signal; Repeat.	POTENTIAL.
M	cell wall, S-layer; Signal; Repeat. <td>POTENTIAL. </td>	POTENTIAL.
T	SIGNAL 1	CELL SURFACE GLYCOPROTEIN 1.
T	CHAIN 29	4 X 156 AA APPROXIMATE REPEATS.
T	DOMAIN 36	
T	REPEAT 36	
T	REPEAT 191	
T	REPEAT 207	
T	REPEAT 409	
T	REPEAT 565	
T	REPEAT 763	
T	REPEAT 607	
T	DOMAIN 771	
T	DOMAIN 1377	
T	1378	
T	1449	
T	DOMAIN 1453	
T	1494	
T	1565	
T	DOMAIN 1495	
T	1565	
T	DOMAIN 1566	
T	1625	
T	DOMAIN 1626	
T	1646	
T	SEQUENCE 1664 AA; 178194 MW; 5f396695b8a9ef74b CRC64;	
Q		

Query Match 18.8%; Score 942; DB 1; Length 1664;
 Best Local Similarity 33.0%; Pred. No. 2, 2e-30;
 Matches 276; Conservative 83; Mismatches 312; Indels 166; Gaps 43

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: L21998; AAB95295.1; -
DR EMBL: M74027; AAB59875.1; -
DR EMBL: M94131; AAB59163.1; -
DR EMBL: M94132; AAB59164.1; -
DR MIM: 158370; -
DR InterPro: IPR000359; Cys_knot.
DR InterPro: IPR000561; EGF_knot.
DR InterPro: IPR002400; GF_cys_knot.
DR InterPro: IPR001007; VMEC.
DR Pfam: PF00007; Cys_knot. 1.
DR Pfam: PF00094; Vwd. 4.
DR PRINTS: PR00438; GRCYSKNOT.
DR SMART: SM00214; VMC. 2.
DR SMART: SM00011; VMC_def. 2.
DR PROSITE: PS00022; EGF_1; UNKNOWN_1.
DR PROSITE: PS01185; CTCK_1; 1.
DR PROSITE: PS01225; CTCK_2; 1.
DR PROSITE: PS01208; VMEC. 2.
DR Glycoprotein: Repeat; Signal.
FT SIGNAL 1 20
FT CHAIN 21 5179
FT DOMAIN 1401 1747
FT REPEAT 1401 1416
FT REPEAT 1417 1432
FT REPEAT 1433 1448
FT REPEAT 1449 1464
FT REPEAT 1465 1471
FT REPEAT 1472 1478
FT REPEAT 1479 1494
FT REPEAT 1495 1517
FT REPEAT 1518 1533
FT REPEAT 1534 1556
FT REPEAT 1557 1572
FT REPEAT 1573 1596
FT REPEAT 1597 1612
FT REPEAT 1613 1635
FT REPEAT 1636 1651
FT REPEAT 1652 1675
FT REPEAT 1676 1693
FT REPEAT 1694 1715
FT REPEAT 1716 1731
FT REPEAT 1732 1747
FT DOMAIN 1748 1763
FT DOMAIN 1764 1779
FT DOMAIN 1780 1795
FT DOMAIN 1796 1811
FT DOMAIN 1812 1827
FT DOMAIN 1828 1843
FT DOMAIN 1844 1859
FT DOMAIN 1860 1875
FT DOMAIN 1876 1891
FT DOMAIN 1892 1907
FT DOMAIN 1908 1923
FT DOMAIN 1924 1939
FT DOMAIN 1940 1955
FT DOMAIN 1956 1971
FT DOMAIN 1972 1987
FT DOMAIN 1988 2003
FT DOMAIN 2004 2019
FT DOMAIN 2020 2035
FT DOMAIN 2036 2051
FT DOMAIN 2052 2067
FT DOMAIN 2068 2083
FT DOMAIN 2084 2099
FT DOMAIN 2100 2115
FT DOMAIN 2116 2131
FT DOMAIN 2132 2147
FT DOMAIN 2148 2163
FT DOMAIN 2164 2179
FT DOMAIN 2180 2195
FT DOMAIN 2196 2211
FT DOMAIN 2212 2227
FT DOMAIN 2228 2243
FT DOMAIN 2244 2259
FT DOMAIN 2260 2275
FT DOMAIN 2276 2291
FT DOMAIN 2292 2307
FT DOMAIN 2308 2323
FT DOMAIN 2324 2339
FT DOMAIN 2340 2355
FT DOMAIN 2356 2371
FT DOMAIN 2372 2387
FT DOMAIN 2388 2403
FT DOMAIN 2404 2419
FT DOMAIN 2420 2435
FT DOMAIN 2436 2451
FT DOMAIN 2452 2467
FT DOMAIN 2468 2483
FT DOMAIN 2484 2499
FT DOMAIN 2500 2515
FT DOMAIN 2516 2531
FT DOMAIN 2532 2547
FT DOMAIN 2548 2563
FT DOMAIN 2564 2579
FT DOMAIN 2580 2595
FT DOMAIN 2596 2611
FT DOMAIN 2612 2627
FT DOMAIN 2628 2643
FT DOMAIN 2644 2659
FT DOMAIN 2660 2675
FT DOMAIN 2676 2691
FT DOMAIN 2692 2707
FT DOMAIN 2708 2723
FT DOMAIN 2724 2739
FT DOMAIN 2740 2755
FT DOMAIN 2756 2771
FT DOMAIN 2772 2787
FT DOMAIN 2788 2803
FT DOMAIN 2804 2819
FT DOMAIN 2820 2835
FT DOMAIN 2836 2851
FT DOMAIN 2852 2867
FT DOMAIN 2868 2883
FT DOMAIN 2884 2899
FT DOMAIN 2900 2915
FT DOMAIN 2916 2931
FT DOMAIN 2932 2947
FT DOMAIN 2948 2963
FT DOMAIN 2964 2979
FT DOMAIN 2980 2995
FT DOMAIN 2996 3011
FT DOMAIN 3012 3027
FT DOMAIN 3028 3043
FT DOMAIN 3044 3059
FT DOMAIN 3060 3075
FT DOMAIN 3076 3091
FT DOMAIN 3092 3107
FT DOMAIN 3108 3123
FT DOMAIN 3124 3139
FT DOMAIN 3140 3155
FT DOMAIN 3156 3171
FT DOMAIN 3172 3187
FT DOMAIN 3188 3203
FT DOMAIN 3204 3219
FT DOMAIN 3220 3235
FT DOMAIN 3236 3251
FT DOMAIN 3252 3267
FT DOMAIN 3268 3283
FT DOMAIN 3284 3299
FT DOMAIN 3300 3315
FT DOMAIN 3316 3331
FT DOMAIN 3332 3347
FT DOMAIN 3348 3363
FT DOMAIN 3364 3379
FT DOMAIN 3380 3395
FT DOMAIN 3396 3411
FT DOMAIN 3412 3427
FT DOMAIN 3428 3443
FT DOMAIN 3444 3459
FT DOMAIN 3460 3475
FT DOMAIN 3476 3491
FT DOMAIN 3492 3507
FT DOMAIN 3508 3523
FT DOMAIN 3524 3539
FT DOMAIN 3540 3555
FT DOMAIN 3556 3571
FT DOMAIN 3572 3587
FT DOMAIN 3588 3603
FT DOMAIN 3604 3619
FT DOMAIN 3620 3635
FT DOMAIN 3636 3651
FT DOMAIN 3652 3667
FT DOMAIN 3668 3683
FT DOMAIN 3684 3699
FT DOMAIN 3700 3715
FT DOMAIN 3716 3731
FT DOMAIN 3732 3747
FT DOMAIN 3748 3763
FT DOMAIN 3764 3779
FT DOMAIN 3780 3795
FT DOMAIN 3796 3811
FT DOMAIN 3812 3827
FT DOMAIN 3828 3843
FT DOMAIN 3844 3859
FT DOMAIN 3860 3875
FT DOMAIN 3876 3891
FT DOMAIN 3892 3907
FT DOMAIN 3908 3923
FT DOMAIN 3924 3939
FT DOMAIN 3940 3955
FT DOMAIN 3956 3971
FT DOMAIN 3972 3987
FT DOMAIN 3988 4003
FT DOMAIN 4004 4019
FT DOMAIN 4020 4035
FT DOMAIN 4036 4051
FT DOMAIN 4052 4067
FT DOMAIN 4068 4083
FT DOMAIN 4084 4099
FT DOMAIN 4100 4115
FT DOMAIN 4116 4131
FT DOMAIN 4132 4147
FT DOMAIN 4148 4163
FT DOMAIN 4164 4179
FT DOMAIN 4180 4195
FT DOMAIN 4196 4211
FT DOMAIN 4212 4227
FT DOMAIN 4228 4243
FT DOMAIN 4244 4259
FT DOMAIN 4260 4275
FT DOMAIN 4276 4291
FT DOMAIN 4292 4307
FT DOMAIN 4308 4323
FT DOMAIN 4324 4339
FT DOMAIN 4340 4355
FT DOMAIN 4356 4371
FT DOMAIN 4372 4387
FT DOMAIN 4388 4403
FT DOMAIN 4404 4419
FT DOMAIN 4420 4435
FT DOMAIN 4436 4451
FT DOMAIN 4452 4467
FT DOMAIN 4468 4483
FT DOMAIN 4484 4499
FT DOMAIN 4500 4515
FT DOMAIN 4516 4531
FT DOMAIN 4532 4547
FT DOMAIN 4548 4563
FT DOMAIN 4564 4579
FT DOMAIN 4580 4595
FT DOMAIN 4596 4611
FT DOMAIN 4612 4627
FT DOMAIN 4628 4643
FT DOMAIN 4644 4659
FT DOMAIN 4660 4675
FT DOMAIN 4676 4691
FT DOMAIN 4692 4707
FT DOMAIN 4708 4723
FT DOMAIN 4724 4739
FT DOMAIN 4740 4755
FT DOMAIN 4756 4771
FT DOMAIN 4772 4787
FT DOMAIN 4788 4803
FT DOMAIN 4804 4819
FT DOMAIN 4820 4835
FT DOMAIN 4836 4851
FT DOMAIN 4852 4867
FT DOMAIN 4868 4883
FT DOMAIN 4884 4899
FT DOMAIN 4900 4915
FT DOMAIN 4916 4931
FT DOMAIN 4932 4947
FT DOMAIN 4948 4963
FT DOMAIN 4964 4979
FT DOMAIN 4980 4995
FT DOMAIN 4996 5011
FT DOMAIN 5012 5027
FT DOMAIN 5028 5043
FT DOMAIN 5044 5059
FT DOMAIN 5060 5075
FT DOMAIN 5076 5091
FT DOMAIN 5092 5107
FT DOMAIN 5108 5123
FT DOMAIN 5124 5139
FT DOMAIN 5140 5155
FT DOMAIN 5156 5171
FT DOMAIN 5172 5187
FT DOMAIN 5188 5203
FT DOMAIN 5204 5219
FT DOMAIN 5220 5235
FT DOMAIN 5236 5251
FT DOMAIN 5252 5267
FT DOMAIN 5268 5283
FT DOMAIN 5284 5299
FT DOMAIN 5300 5315
FT DOMAIN 5316 5331
FT DOMAIN 5332 5347
FT DOMAIN 5348 5363
FT DOMAIN 5364 5379
FT DOMAIN 5380 5395
FT DOMAIN 5396 5411
FT DOMAIN 5412 5427
FT DOMAIN 5428 5443
FT DOMAIN 5444 5459
FT DOMAIN 5460 5475
FT DOMAIN 5476 5491
FT DOMAIN 5492 5507
FT DOMAIN 5508 5523
FT DOMAIN 5524 5539
FT DOMAIN 5540 5555
FT DOMAIN 5556 5571
FT DOMAIN 5572 5587
FT DOMAIN 5588 5603
FT DOMAIN 5604 5619
FT DOMAIN 5620 5635
FT DOMAIN 5636 5651
FT DOMAIN 5652 5667
FT DOMAIN 5668 5683
FT DOMAIN 5684 5699
FT DOMAIN 5700 5715
FT DOMAIN 5716 5731
FT DOMAIN 5732 5747
FT DOMAIN 5748 5763
FT DOMAIN 5764 5779
FT DOMAIN 5780 5795
FT DOMAIN 5796 5811
FT DOMAIN 5812 5827
FT DOMAIN 5828 5843
FT DOMAIN 5844 5859
FT DOMAIN 5860 5875
FT DOMAIN 5876 5891
FT DOMAIN 5892 5907
FT DOMAIN 5908 5923
FT DOMAIN 5924 5939
FT DOMAIN 5940 5955
FT DOMAIN 5956 5971
FT DOMAIN 5972 5987
FT DOMAIN 5988 6003
FT DOMAIN 6004 6019
FT DOMAIN 6020 6035
FT DOMAIN 6036 6051
FT DOMAIN 6052 6067
FT DOMAIN 6068 6083
FT DOMAIN 6084 6099
FT DOMAIN 6100 6115
FT DOMAIN 6116 6131
FT DOMAIN 6132 6147
FT DOMAIN 6148 6163
FT DOMAIN 6164 6179
FT DOMAIN 6180 6195
FT DOMAIN 6196 6211
FT DOMAIN 6212 6227
FT DOMAIN 6228 6243
FT DOMAIN 6244 6259
FT DOMAIN 6260 6275
FT DOMAIN 6276 6291
FT DOMAIN 6292 6307
FT DOMAIN 6308 6323
FT DOMAIN 6324 6339
FT DOMAIN 6340 6355
FT DOMAIN 6356 6371
FT DOMAIN 6372 6387
FT DOMAIN 6388 6403
FT DOMAIN 6404 6419
FT DOMAIN 6420 6435
FT DOMAIN 6436 6451
FT DOMAIN 6452 6467
FT DOMAIN 6468 6483
FT DOMAIN 6484 6499
FT DOMAIN 6500 6515
FT DOMAIN 6516 6531
FT DOMAIN 6532 6547
FT DOMAIN 6548 6563
FT DOMAIN 6564 6579
FT DOMAIN 6580 6595
FT DOMAIN 6596 6611
FT DOMAIN 6612 6627
FT DOMAIN 6628 6643
FT DOMAIN 6644 6659
FT DOMAIN 6660 6675
FT DOMAIN 6676 6691
FT DOMAIN 6692 6707
FT DOMAIN 6708 6723
FT DOMAIN 6724 6739
FT DOMAIN 6740 6755
FT DOMAIN 6756 6771
FT DOMAIN 6772 6787
FT DOMAIN 6788 6803
FT DOMAIN 6804 6819
FT DOMAIN 6820 6835
FT DOMAIN 6836 6851
FT DOMAIN 6852 6867
FT DOMAIN 6868 6883
FT DOMAIN 6884 6899
FT DOMAIN 6900 6915
FT DOMAIN 6916 6931
FT DOMAIN 6932 6947
FT DOMAIN 6948 6963
FT DOMAIN 6964 6979
FT DOMAIN 6980 6995
FT DOMAIN 6996 7011
FT DOMAIN 7012 7027
FT DOMAIN 7028 7043
FT DOMAIN 7044 7059
FT DOMAIN 7060 7075
FT DOMAIN 7076 7091
FT DOMAIN 7092 7107
FT DOMAIN 7108 7123
FT DOMAIN 7124 7139
FT DOMAIN 7140 7155
FT DOMAIN 7156 7171
FT DOMAIN 7172 7187
FT DOMAIN 7188 7203
FT DOMAIN 7204 7219
FT DOMAIN 7220 7235
FT DOMAIN 7236 7251
FT DOMAIN 7252 7267
FT DOMAIN 7268 7283
FT DOMAIN 7284 7299
FT DOMAIN 7300 7315
FT DOMAIN 7316 7331
FT DOMAIN 7332 7347
FT DOMAIN 7348 7363
FT DOMAIN 7364 7379
FT DOMAIN 7380 7395
FT DOMAIN 7396 7411
FT DOMAIN 7412 7427
FT DOMAIN 7428 7443
FT DOMAIN 7444 7459
FT DOMAIN 7460 7475
FT DOMAIN 7476 7491
FT DOMAIN 7492 7507
FT DOMAIN 7508 7523
FT DOMAIN 7524 7539
FT DOMAIN 7540 7555
FT DOMAIN 7556 7571
FT DOMAIN 7572 7587
FT DOMAIN 7588 7603
FT DOMAIN 7604 7619
FT DOMAIN 7620 7635
FT DOMAIN 7636 7651
FT DOMAIN 7652 7667
FT DOMAIN 7668 7683
FT DOMAIN 7684 7699
FT DOMAIN 7700 7715
FT DOMAIN 7716 7731
FT DOMAIN 7732 7747
FT DOMAIN 7748 7763
FT DOMAIN 7764 7779
FT DOMAIN 7780 7795
FT DOMAIN 7796 7811
FT DOMAIN 7812 7827
FT DOMAIN 7828 7843
FT DOMAIN 7844 7859
FT DOMAIN 7860 7875
FT DOMAIN 7876 7891
FT DOMAIN 7892 7907
FT DOMAIN 7908 7923
FT DOMAIN 7924 7939
FT DOMAIN 7940 7955
FT DOMAIN 7956 7971
FT DOMAIN 7972 7987
FT DOMAIN 7988 8003
FT DOMAIN 8004 8019
FT DOMAIN 8020 8035
FT DOMAIN 8036 8051
FT DOMAIN 8052 8067
FT DOMAIN 8068 8083
FT DOMAIN 8084 8099
FT DOMAIN 8100 8115
FT DOMAIN 8116 8131
FT DOMAIN 8132 8147
FT DOMAIN 8148 8163
FT DOMAIN 8164 8179
FT DOMAIN 8180 8195
FT DOMAIN 8196 8211
FT DOMAIN 8212 8227
FT DOMAIN 8228 8243
FT DOMAIN 8244 8259
FT DOMAIN 8260 8275
FT DOMAIN 8276 8291
FT DOMAIN 8292 8307
FT DOMAIN 8308 8323
FT DOMAIN 8324 8339
FT DOMAIN 8340 8355
FT DOMAIN 8356 8371
FT DOMAIN 8372 8387
FT DOMAIN 8388 8403
FT DOMAIN 8404 8419
FT DOMAIN 8420 8435
FT DOMAIN 8436 8451
FT DOMAIN 8452 8467
FT DOMAIN 8468 8483
FT DOMAIN 8484 8499
FT DOMAIN 8500 8515
FT DOMAIN 8516 8531
FT DOMAIN 8532 8547
FT DOMAIN 8548 8563
FT DOMAIN 8564 8579
FT DOMAIN 8580 8595
FT DOMAIN 8596 8611
FT DOMAIN 8612 8627
FT DOMAIN 8628 8643
FT DOMAIN 8644 8659
FT DOMAIN 8660 8675
FT DOMAIN 8676 8691
FT DOMAIN 8692 8707
FT DOMAIN 8708 8723
FT DOMAIN 8724 8739
FT DOMAIN 8740 8755
FT DOMAIN 8756 8771
FT DOMAIN 8772 8787
FT DOMAIN 8788 8803
FT DOMAIN 8804 8819
FT DOMAIN 8820 8835
FT DOMAIN 8836 8851
FT DOMAIN 8852 8867
FT DOMAIN 8868 8883
FT DOMAIN 8884 8899
FT DOMAIN 8900 8915
FT DOMAIN 8916 8931
FT DOMAIN 8932 8947
FT DOMAIN 8948 8963
FT DOMAIN 8964 8979
FT DOMAIN 8980 8995
FT DOMAIN 8996 9011
FT DOMAIN 9012 9027
FT DOMAIN 9028 9043
FT DOMAIN 9044 9059
FT DOMAIN 9060 9075
FT DOMAIN 9076 9091
FT DOMAIN 9092 9107
FT DOMAIN 9108 9123
FT DOMAIN 9124 9139
FT DOMAIN 9140 9155
FT DOMAIN 9156 9171
FT DOMAIN 9172 9187
FT DOMAIN 9188 9203
FT DOMAIN 9204 9219
FT DOMAIN 9220 9235
FT DOMAIN 9236 9251
FT DOMAIN 9252 9267
FT DOMAIN 9268 9283
FT DOMAIN 9284 9299
FT DOMAIN 9300 9315
FT DOMAIN 9316 9331
FT DOMAIN 9332 9347
FT DOMAIN 9348 9363
FT DOMAIN 9364 9379
FT DOMAIN 9380 9395
FT DOMAIN 9396 9411
FT DOMAIN 9412 9427
FT DOMAIN 9428 9443
FT DOMAIN 9444 9459
FT DOMAIN 9460 9475
FT DOMAIN 9476 9491
FT DOMAIN 9492 9507
FT DOMAIN 9508 9523
FT DOMAIN 9524 9539
FT DOMAIN 9540 9555
FT DOMAIN 9556 9571
FT DOMAIN 9572 9587
FT DOMAIN 9588 9603
FT DOMAIN 9604 9619
FT DOMAIN 9620 9635
FT DOMAIN 9636 9651
FT DOMAIN 9652 9667
FT DOMAIN 9668 9683
FT DOMAIN 9684 9699
FT DOMAIN 9700 9715
FT DOMAIN 9716 9731
FT DOMAIN 9732 9747
FT DOMAIN 9748 9763
FT DOMAIN 9764 9779
FT DOMAIN 9780 9795
FT DOMAIN 9796 9811
FT DOMAIN 9812 9827
FT DOMAIN 9828 9843
FT DOMAIN 9844 9859
FT DOMAIN 9860 9875
FT DOMAIN 9876 9891
FT DOMAIN 9892 9907
FT DOMAIN 9908 9923
FT DOMAIN 9924 9939
FT DOMAIN 9940 9955
FT DOMAIN 9956 9971
FT DOMAIN 9972 9987
FT DOMAIN 9988 10003
FT DOMAIN 10004 10019
FT DOMAIN 10020 10035
FT DOMAIN 10036 10051
FT DOMAIN 10052 10067
FT DOMAIN 10068 10083
FT DOMAIN 10084 10099
FT DOMAIN 10100 10115
FT DOMAIN 10116 10131
FT DOMAIN 10132 10147
FT DOMAIN 10148 10163
FT DOMAIN 10164 10179
FT DOMAIN 10180 10195
FT DOMAIN 10196 10211
FT DOMAIN 10212 10227
FT DOMAIN 10228 10243
FT DOMAIN 10244 10259
FT DOMAIN 10260 10275
FT DOMAIN 10276 10291
FT DOMAIN 10292 10307
FT DOMAIN 10308 10323
FT DOMAIN 10324 10339
FT DOMAIN 10340 10355
FT DOMAIN 10356 10371
FT DOMAIN 10372 10387
FT DOMAIN 10388 10403
FT DOMAIN 10404 10419
FT DOMAIN 10420 10435
FT DOMAIN 10436 10451
FT DOMAIN 10452 10467
FT DOMAIN 10468 10483
FT DOMAIN 10484 10499
FT DOMAIN 10500 10515
FT DOMAIN 10516 10531
FT DOMAIN 10532 10547
FT DOMAIN 10548 10563
FT DOMAIN 10564 10579
FT DOMAIN 10580 10595
FT DOMAIN 10596 10611
FT DOMAIN 10612 10627
FT DOMAIN 10628 10643
FT DOMAIN 10644 10659
FT DOMAIN 10660 10675
FT DOMAIN 10676 10691
FT DOMAIN 10692 10707
FT DOMAIN 10708 10723
FT DOMAIN 10724 10739
FT DOMAIN 10740 10755
FT DOMAIN 10756 10771
FT DOMAIN 10772 10787
FT DOMAIN 10788 10803
FT DOMAIN 10804 10819
FT DOMAIN 10820 10835
FT DOMAIN 10836 10851
FT DOMAIN 10852 10867
FT DOMAIN 10868 10883
FT DOMAIN 10884 10899
FT DOMAIN 10900 10915
FT DOMAIN 10916 10931
FT DOMAIN 10932 10947
FT DOMAIN 10948 10963
FT DOMAIN 10964 10979
FT DOMAIN 10980 10995
FT DOMAIN 10996 11011
FT DOMAIN 11012 11027
FT DOMAIN 11028 11043
FT DOMAIN 11044 11059
FT DOMAIN 11060 11075
FT DOMAIN 11076 11091
FT DOMAIN 11092 11107
FT DOMAIN 11108 11123
FT DOMAIN 11124 11139
FT DOMAIN 11140 11155
FT DOMAIN 11156 11171
FT DOMAIN 11172 11187
FT DOMAIN 11188 11203
FT DOMAIN 11204 11219
FT DOMAIN 11220 11235
FT DOMAIN 11236 11251
FT DOMAIN 11252 11267
FT DOMAIN 11268 11283
FT DOMAIN 11284 11299
FT DOMAIN 11300 11315
FT DOMAIN 11316 11331
FT DOMAIN 11332 11347
FT DOMAIN 11348 11363
FT DOMAIN 11364 11379
FT DOMAIN 11380 11395
FT DOMAIN 11396 11411
FT DOMAIN 11412 11427
FT DOMAIN 11428 11443
FT DOMAIN 11444 11459
FT DOMAIN 11460 11475
FT DOMAIN 11476 11491
FT DOMAIN 11492 11507
FT DOMAIN 11508 11523
FT DOMAIN 11524 11539
FT DOMAIN 11540 11555
FT DOMAIN 11556 11571
FT DOMAIN 11572 11587
FT DOMAIN 11588 11603
FT DOMAIN 11604 11619
FT DOMAIN 11620 11635
FT DOMAIN 11636 11651
FT DOMAIN 11652 11667
FT DOMAIN 11668 11683
FT DOMAIN 11684 11699
FT DOMAIN 11700 11715
FT DOMAIN 11716 11731
FT DOMAIN 11732 11747
FT DOMAIN 11748 11763
FT DOMAIN 11764 11779
FT DOMAIN 11780 11795
FT DOMAIN 11796 11811
FT DOMAIN 11812 11827
FT DOMAIN 11828 11843
FT DOMAIN 11844 11859
FT DOMAIN 11860 11875
FT DOMAIN 11876 11891
FT DOMAIN 11892 11907
FT DOMAIN 11908 11923
FT DOMAIN 11924 11939
FT DOMAIN 11940 11955
FT DOMAIN 11956 11971
FT DOMAIN 11972 11987
FT DOMAIN 11988 12003
FT DOMAIN 12004 12019
FT DOMAIN 12020 12035
FT DOMAIN 12036 12051
FT DOMAIN 12052 12067
FT DOMAIN 12068 12083
FT DOMAIN 12084 12099
FT DOMAIN 12100 12115
FT DOMAIN 12116 12131
FT DOMAIN 12132 12147
FT DOMAIN 12148 12163
FT DOMAIN 12164 12179
FT DOMAIN 12180 12195
FT DOMAIN 12196 12211
FT DOMAIN 12212 12227
FT DOMAIN 12228 12243
FT DOMAIN 12244 12259
FT DOMAIN 12260 12275
FT DOMAIN 12276 12291
FT DOMAIN 12292 12307
FT DOMAIN 12308 12323
FT DOMAIN 12324 12339
FT DOMAIN 12340 12355
FT DOMAIN 12356 12371
FT DOMAIN 12372 12387
FT DOMAIN 12388 12403
FT DOMAIN 12404 12419
FT DOMAIN 12420 12435
FT DOMAIN 12436 12451
FT DOMAIN 12452 12467
FT DOMAIN 12468 12483
FT DOMAIN 12484 12499
FT DOMAIN 12500 12515
FT DOMAIN 12516 12531
FT DOMAIN 12532 12547
FT DOMAIN 12548 12563
FT DOMAIN 12564 12579
FT DOMAIN 12580 12595
FT DOMAIN 12596 12611
FT DOMAIN 12612 12627
FT DOMAIN 12628 12643
FT DOMAIN 12644 12659
FT DOMAIN 12660 12675
FT DOMAIN 12676 12691
FT DOMAIN 12692 12707
FT DOMAIN 12708 12723
FT DOMAIN 12724 12739
FT DOMAIN 12740 12755
FT DOMAIN 12756 12771
FT DOMAIN 12772 12787
FT DOMAIN 12788 12803
FT DOMAIN 12804 12819
FT DOMAIN 12820 12835
FT DOMAIN 12836 12851
FT DOMAIN 12852 12867
FT DOMAIN 12868 12883
FT DOMAIN 12884 12899
FT DOMAIN 12900 12915
FT DOMAIN 12916 12931
FT DOMAIN 12932 12947
FT DOMAIN 12948 12963
FT DOMAIN 12964 12979
FT DOMAIN 12980 12995
FT DOMAIN 12996 13011
FT DOMAIN 13012 13027
FT DOMAIN 13028 13043
FT DOMAIN 13044 13059
FT DOMAIN 13060 13075
FT DOMAIN 13076 13091
FT DOMAIN 13092 13107
FT DOMAIN 13108 13123
FT DOMAIN 13124 13139
FT DOMAIN 13140 13155
FT DOMAIN 13156 13171
FT DOMAIN 13172 13187
FT DOMAIN 13188 13203
FT DOMAIN 13204 13219
FT DOMAIN 13220 13235
FT DOMAIN 13236 13251
FT DOMAIN 13252 13267
FT DOMAIN 13268 13283
FT DOMAIN 13284 13299
FT DOMAIN 13300 13315
FT DOMAIN 13316 13331
FT DOMAIN 13332 13347
FT DOMAIN 13348 13363
FT DOMAIN 13364 13379
FT DOMAIN 13380 13395
FT DOMAIN 13396 13411
FT DOMAIN 13412 13427
FT DOMAIN 13428 13443
FT DOMAIN 13444 13459
FT DOMAIN 13460 13475
FT DOMAIN 13476 13491
FT DOMAIN 13492 13507
FT DOMAIN 13508 13523
FT DOMAIN 13524 13539
FT DOMAIN 13540 13555
FT DOMAIN 13556 13571
FT DOMAIN 13572 13587
FT DOMAIN 13588 13603
FT DOMAIN 13604 13619
FT DOMAIN 13620 13635
FT DOMAIN 13636 13651
FT DOMAIN 13652 13667
FT DOMAIN 13668 13683
FT DOMAIN 13684 13699
FT DOMAIN 13700 13715
FT DOMAIN 13716 13731
FT DOMAIN 13732 13747
FT DOMAIN 13748 13763
FT DOMAIN 13764 13779
FT DOMAIN 13780 13795
FT DOMAIN 13796 13811
FT DOMAIN 13812 13827
FT DOMAIN 13828 13843
FT DOMAIN 13844 13859
FT DOMAIN 13860 13875
FT DOMAIN 13876 13891
FT DOMAIN 13892 13907
FT DOMAIN 13908 13923
FT DOMAIN 13924 13939
FT DOMAIN 13940 13955
FT DOMAIN 13956 13971
FT DOMAIN 13972 13987
FT DOMAIN 13988 14003
FT DOMAIN 14004 14019
FT DOMAIN 14020 14035
FT DOMAIN 14036 14051
FT DOMAIN 14052 14067
FT DOMAIN 14068 14083

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:13:35 ; Search time 47.4 Seconds
(without alignments)
727.882 Million cell updates/sec

Title: US-09-556-246-1_COPY_200_1140
5011
Perfect score: 1 VKONKKNRRTKKRPKPPV.....DMDYLPVPNGIINPMLS 941
Sequence:

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_39.*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1168	23.3	5179	1 MUC2_HUMAN	Q02817 homo sapien
2	942	18.8	1664	1 SLPL_CLOTM	Q06852 clostridium
3	762.3	15.2	1367	1 AMYH_YEAST	P08640 saccharomyc
4	651	13.0	1255	1 MUC1_HUMAN	P15941 h mucin 1 p
5	556.5	11.1	875	1 FPL_MYTD	Q25460 mytilus edu
6	555.5	11.1	2700	1 ZAN_HUMAN	Q9Y493 homo sapien
7	551	11.0	620	1 EXTN_TOBAC	P13963 nicotiana t
8	533	10.6	1087	1 NEH_MOUSE	P19246 mus musculu
9	530.5	10.6	1162	1 TCNA_TRYCR	P23253 trypanosoma
10	530	10.6	865	1 CPN_DROME	Q02910 drosophila
11	518.5	10.3	872	1 PPL_MYCO	Q25434 mytilus cor
12	509.5	10.2	662	1 MUC1_XENLA	Q05049 xenopus lae
13	503.5	10.0	1970	1 RPB1_HUMAN	P24928 homo sapien
14	498.5	9.9	1970	1 RPB1_MOUSE	P08775 mus musculu
15	497	9.9	831	1 NEH_RAT	P16884 rattus norv
16	493.5	9.8	467	1 RPB1_CRIGR	P11414 placetulus
17	488.5	9.7	826	1 SSP2_PLAYO	O01443 placetulus
18	488	9.7	1020	1 NEH_HUMAN	P12036 homo sapien
19	475.5	9.5	267	1 EXTN_MAIZE	P46599 candida alb
20	468.5	9.3	634	1 HNP1_CANAL	O08799 mus musculu
21	467.5	9.3	5376	1 ZAN_MOUSE	P31370 saccharomyc
22	467	9.3	817	1 VRP1_YEAST	P47179 saccharomyc
23	454	9.1	1161	1 VJ9P_YEAST	P48634 homo sapien
24	452.5	9.0	2142	1 BAT2_HUMAN	O00130 ictalurid h
25	448.5	9.0	670	1 VG50_HSV1	P28968 equine herp
26	448	8.9	797	1 VGIX_HSVB	Q27409 mytilus gal
27	442.5	8.8	751	1 FPL_MYTGA	O00268 homo sapien
28	439.5	8.8	1083	1 T2D3_HUMAN	P17437 xenopus lae
29	432.5	8.6	439	1 XP2_XENLA	P10220 herpes simp
30	432	8.5	3164	1 TEGU_HSV1	O9um66 homo sapien
31	426	8.5	2715	1 TRX2_HUMAN	P27546 mus musculu
32	424.5	8.5	1125	1 MAP4_MOUSE	P02840 drosophila
33	424	8.5	307	1 S6S3_DROME	

34	424	8.5	1251	1 Y0U3_CAEEL	O09550 caenorhabdi
35	422.5	8.4	1794	1 YAVI_SCHPO	O10172 schizosacch
36	422.5	8.4	3421	1 TEGU_HSVB	P28955 equine herp
37	419.5	8.4	2476	1 ZAN_PTG	O28983 sus scrofa
38	419	8.4	1185	1 DRPL_HUMAN	P54259 homo sapien
39	415	8.3	1229	1 N121_HUMAN	O9Y2N3 homo sapien
40	411.5	8.2	907	1 VGP3_EBV	P03200 Epstein-bar
41	410	8.2	1183	1 DRPL_RAT	P54258 rattus norv
42	406	8.1	2774	1 MAP4_RAT	P34926 rattus norv
43	406	8.1	3256	1 K167_HUMAN	P46013 homo sapien
44	405	8.1	1411	1 TCOF_HUMAN	O13428 homo sapien
45	400.5	8.0	2517	1 NCR2_HUMAN	O9Y618 h nuclear r

ALIGNMENTS

RESULT	1	STANDARD:	PRT:	5179 AA.
MUC2_HUMAN				
ID	Q02817; Q14878;			
AC	01-JUN-1994 (Rel. 29, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	MUCIN 2 PRECURSOR (INTESTINAL MUCIN 2).			
GN	MUC2 OR SMC.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Intestine;			
RX	MEDLINE=94132002; PubMed=8300571;			
RA	Gum J.R. Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;			
RT	"Molecular cloning of human intestinal mucin (MUC2) cDNA.			
RT	Identification of the amino terminus and overall sequence similarity			
RT	to prepro-von Willebrand factor.";			
RL	J. Biol. Chem. 269:2440-2446(1994).			
RN	[2]			
RP	SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.			
RC	TISSUE=Colon;			
RX	MEDLINE=93016075; PubMed=1400449;			
RA	Gum J.R. Jr., Hicks J.W., Toribara N.W., Rothe E.-M., Lagace R.E.,			
RA	Kim Y.S.;			
RT	"The human MUC2 intestinal mucin has cysteine-rich subdomains located			
RT	both upstream and downstream of its central repetitive region.";			
RL	J. Biol. Chem. 267:21375-21383(1992).			
RN	[3]			
RP	SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.			
RX	MEDLINE=91358717; PubMed=1855763;			
RA	Toribara N.W., Gum J.R. Jr., Culhane P.J., Lagace R.E., Hicks J.W.,			
RA	Petersen G.M., Kim Y.S.;			
RT	"MUC-2 human small intestinal mucin gene structure. Repeated arrays			
RT	and polymorphism.";			
RL	J. Clin. Invest. 88:1005-1013(1991).			
CC	-1- FUNCTION: CONS. THE EPITHELIA OF THE INTESTINES, AIRWAYS, AND			
CC	OTHER MUCUS MEMBRANE-CONTAINING ORGANS. THOUGHT TO PROVIDE A			
CC	PROTECTIVE, LUBRICATING BARRIER AGAINST PARTICLES AND INFECTION			
CC	AGENTS AT MUCOSAL SURFACES.			
CC	-1- SUBUNIT: MULTIMERIC.			
CC	-1- SUBCELLULAR LOCATION: SECRETED.			
CC	-1- TISSUE SPECIFICITY: COLON, SMALL INTESTINE, COLONIC TUMORS,			
CC	BROCHUS, CERVIX AND GALL BLADDER.			
CC	-1- PPM: ALL Cysteine residues are INVOLVED IN INTRACHAIN OR			
CC	INTERCHAIN DISULFIDE BONDS (BY SIMILARITY).			
CC	-1- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC AND			
CC	VARIES AMONG DIFFERENT ALLELES.			
CC	-1- SIMILARITY: THE N-TERMINAL DOMAIN SHOWS SOME SIMILARITY TO THAT			
CC	OF SLIKWORM HEMOCYTIN.			
CC	-1- SIMILARITY: CONTAINS 2 WPC DOMAINS.			
CC	-1- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).			

```
QY 117 TSAKDLAPTSKVLAKPTPKAETTTGPAULTTPEKEPTTPTKE-----PASTTPEKEPTPTTI 172
Db 171 EASKKKKQVIAV----PKQAVVAKGLASNNCKTADSSSESDSPAKKTAATPTPT- 225
QY 173 KSAPTTPEKEPAPTTTTSAPTTPKEPAPTTTKEPAPTTPEKEPAPTTT--TKEPAPTT-----T 226
Db 226 --KPATAAKPOAKKTAGKSSKSSREDSDEQKTAKSKPKPDVYSAVPPPTSVSKKT 283
QY 227 KSAPTTPEKEPAPTT-----TPKKPAPTTPEKEPAPTTT---KEPTPTTPEKEPAPTT 272
Db 284 LSQPGTKAKPSSSDSSDSEBOPAKKATVPAAKAAASAPKAPLAKAETSTDESDSS 343
QY 273 KEPAPTTPEKEPAPTAAPKAPAPTTPEKEPAPTTT---KEPAP-----TTP 311
Db 344 EDEKSSSVKLGVAAPK--APAAP--DAKSTPVAAKKAPAKKASSSDSDSSNEETT 400
QY 312 TKEPSPTTPEKEPAPTTTTSAPTTPKEPAPTTTKEPAPTTKSAPTTPKEPSPTTPEKEPAPTTKEPA- 370
Db 401 TKPAKTTTPAKSAATPTSKTPTNGK---ATPTSKTPAKPGTPTSTAKKSDSSSDSS 457
QY 371 ---PTTPKKPAPTTPEKEPAPT--TPKEPAPTTTTPKAPAPKAPKA-PTTPKETA----- 418
Db 458 SDEETTTTPAKAKTTTPAKSAATPTSKTP---TNSKATPTSKTTPAKPGTPTSAKKDSS 514
QY 419 -----PTTPKKLTPTTPKLAPTTPEKAPAPTTPEELAPTTPEEPTPTTPEEAPPTP--KA 472
Db 515 SDSSSDSSDEKTPA--KRAAKTTTPAKPA-----AKTTTPAKP-----AKTTTPAKP 558
QY 473 AAPNTPEKEPAPT---TPKEPAPTTTPEKEPAPTTPEKETAPTTTP--KGTAPTTTKEPAPTTPK 527
Db 559 AKSTPGKQVPTKSSSDSSSDSSDEKKSAPKPAVKTTPGKATG-----KPVVAS 611
QY 528 KPAKELAPTTTKEPTSTTSDKPAPTTPKGTAPTTPEKEPAPTT--PKEPAPTTPKGTAPT 585
Db 612 KPVPK-----KASSSDSDSSSEETTKTKPLTLSPAVKTLPPKKAESSSDSSSDSD 665
QY 586 TLKEPAPTTPEKAPKELAPTTTGP-----TSTSDKPAPTTPKETAPT--TPKEP 635
Db 666 SEK-----TKPAKPPAKSATPVNTKAPAKONKASKASCSDSSSEEGKSQPTGKSPA 722
QY 636 APPTPKK--PAPTTPEPTPEPTTSEVSTPTTKE--PTTIHKSPESTPELSAEPPTKALENS 693
Db 723 ATAPPKKNPVAVNKDKSSSSSDSGDDEKQKQAAAAKADVQAKAAKPTPKKAASS 782
QY 694 PKEGVPPTTKTPAATKPEMTTAKDKTTERDLRTTPETTTAAPKMTKETATTEKTESK 753
Db 783 SSE---DSSDEDEVSKAKKTWTAVSKSPV---TTPKAVPAKAKSSSESSESDEKQGG 835
QY 754 ITATTOVTSTTODTTPKITTILKTTTLAKVYTTK--KTITTEIMNKPEETA--KPKD 810
Db 836 KNTSTTKIANST-----PKAAAECSSESSSSEDEGKANGTSGKRR 877
QY 811 RATNS---KATTPPKQKPTKAPKKTSTKPKPTMPRVKPKPTTPTP--RKMTSTWPELNPT 866
Db 878 ESTGNACEAVTPE-----NKKLAKASPNTPKYNKKEKLNTPFRRYVEEDIEINP- 928
QY 867 SRIAE 871
Db 929 -RMAD 932
```

Search completed: April 26, 2002, 16:18:10
Job time: 280 sec

A:Accession: A26877
A:Molecule type: DNA
A:Residues: 1-242 <YAM>
A:Cross-references: EMBL:M16164; NID:g172522; P1DN:AAA5014.1; PID:g172525
A:Accession: B26877
A:Molecule type: DNA
A:Residues: 762-1331 <YAZ>
A:Cross-references: EMBL:M16165; NID:g172523; P1DN:AAA5015.1; PID:g172526
R.Pardo, J.M.; Ianez, E.; Zalacain, M.; Claros, M.G.; Jimenez, A.
FBS Lett. 239, 179-184, 1988
A:Title: Similar short elements in the 5' regions of the STA2 and SCA genes from Sacchar
A:Reference number: S27281; MUID:89031230
A:Accession: S27281
A:Molecule type: DNA
A:Residues: 1-31 <PAR>
A:Cross-references: EMBL:X13857; NID:94551; P1DN:CAA32069.1; PID:g4552
R.Lambrecht, M.G.; Bauer, F.F.; Marmur, J.; Pretorius, I.S.
Proc. Natl. Acad. Sci. U.S.A. 93, 8419-8424, 1996
A:Title: Muc1, a mucin-like protein that is regulated by Msi10, is critical for pseudoty
A:Accession: J06123
A:Reference number: J06123; MUID:96323237
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-1367 <LAM>
A:Cross-references: GB:U30626; NID:g1304386; P1DN:AAC49609.1; PID:g1304387
A:Accession: G1304387
A:Gene: SGD:MUC1; STA2; MAL5; DEX2; SGD:S0001458
A:Cross-references: MIPS:YIK019C; SGD:S0001458
A:Map position: 9K
C:Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase
C:Keywords: glycosidase; hydrolase; polysaccharide degradation; transmembrane protein
F:5-21/Domain: transmembrane #status predicted <TM>
F:1350-1366/Domain: transmembrane #status predicted <TM>

Query Match 15.2%; Score 762.5; DB 1; Length 1367;
Best Local Similarity 27.8%; Pred. No. 1.7e-25;
Matches 280; Conservative 115; Mismatches 441; Indels 171; Gaps 42;

```
QY 35 TTPTSTTQHNKVTSPKITAKINP-----RSLPNSDTSKETSJLVNKETTYE 86
D 236 STSSSTSTSSSTSSSTSSSTTAPATPTTSSCKEKPTPTTSSCKEKPTPHNDHTPC 295
QY 87 TKETTTNNKOTSDGKERTSAKETOSIEKTSADLPATSKVLAKPRKAEET--TKGRA 144
D 296 TTKTTTTSK-TCT--KTTTTPVTPS--SSTTESSAPV-----PSSSTTESSAPV 344
QY 145 LTPKE-----PPTTPKEPASTTPEPTTIKSA-----TPKEPAPTTSAPTP 194
D 345 TSSSTESSAPVPTSSSTTESSAPVTSSTTESSAPVTSSTTESSAPVPTSSSTTE 404
QY 195 KEPA-----TTKEPA-----TTKEPA-----TTKSAPTKEPA 238
D 405 SSSAPVTSSTTESSAPVTSSTTESSAPVTSSTTESSAPVTSSTTESSAPVPTSS 464
QY 239 TTPKKAPPTTPKAPPTTPKEPTP-----TPKEPAPTKE-----PAPT--TPKEPA 284
D 465 TTSSSAPVTSSTTESSAPVPTSSSTTESSAPVTSSTTESSAPVPTSSSTTESS 524
QY 285 PTPKAPPTTPKEPAPTKE-----PAPTTPKESPTTPKAPADPTTTSAPPTTK 336
D 525 APAPVTSSTTESSAPVTSSTTESSAPVPTSSSTTSSSTTPVTSSTTESSAPVPT 583
QY 337 EPAPTTKAPPTTPKEPAPTTPKEPAPTTPKEPAPTTPKAPPT-----PKEPAPTTP 390
D 584 -PSSSTTESSAPVPTSSSTTESSAPAPVPTSSSTTESSAPVTSSTTESSAPVPT-- 640
QY 391 KEPAPTTTPKAPATKAPATTPKETAPPTPKLITPTPEKLA-----TPKEKAPTTP 446
D 641 --PSSSTTESSAPVPTSSSTTESSAP--VPTSSSTTESSAPVTSSTTESSAPVTS 697
QY 447 EELAPTPPEPTP--TTPEEPAPT--TPKAAPNTPKEPA-----TPKEPAPT-- 493
D 493
```

```
D 698 STTESSAPVPTSSSTTESSAPVPTSSSTTESSAPVPTSSSTTESSAPVTSST 757
QY 494 -----PKEPAPTTPKETAPT--TPKGAPTTLKEPAPTTPKAPKAPLELAPTTPKEPTST 545
D 758 ESSAPVPTSSSTTESSAPVPTSSSTTESSAPVPTSSSTTESSAPVPTSSSSN 817
QY 546 TSDKAPPTTPKGLAPPTTPKEPAPTTPKEPAPTPKGNAPTTLKEPAPTTPKAPKELAP 605
D 818 ITSSAPSTPEPSSSTTESSVVP-----PSSSTTESSAPVSSSTTESSVAPVPTSSSSN 874
QY 606 TTKGP-----TSTSDKAPAPTTPKETAPPTTPKEPAPTTPKAPPTTP--PPTTSEV 658
D 875 ITSSAPSTPEPSSSTTESSSTGTT--VTPSSSKTPGSGTSTSVSTETTVTPKPTTV 931
QY 659 STPTTKEPTTIHKSPESTPELSAEPPTPKALENSKPEPVPPTTPKPAATKPEPTT-- 714
D 932 TTPSTTTTITVTCSTGNSAGETTSKCPKTTT-----VPTTTTTSVTSTTTTIT 986
QY 715 ---TAKDTERDLRTPET--TAAKPMKETATTEKTESKITAATTOVST--TQD 768
D 987 VCSGTNSAGETTSKCPKTTTTPCSTSPSEASESTTSPPTTVTVSTTVTTT 1046
QY 769 TTPPK-----ITLKTTLA--PKYT--TKTITTEI--NKPDEET-- 805
D 1047 STSKTPGGEITTFVTKNIPPTVLTTLIAPVSVTVTNTPTTTTITVTCSTGNSAGET 1106
QY 806 --AKPKRATNSKATTPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPK 863
D 1107 SCSPKTVTTTVPSCSTGTEGTTTATVTTTAVTTTAVTTTAVTTTAVTTTAVTTTAV 1166
QY 864 NPTSRIAEAMLOTTTRPNO--TPNSKLYEVNPKSED--AGAGEGT 905
D 1167 VPTTV-----TTLAPSAVPTATNAVPTTTTITTECSAATNAGET 1207
```

RESULT 8
T18535
high molecular mass nuclear antigen - chicken (fragment)
C:Species: Gallus gallus (chicken)
C:date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T18535
J.Shimada, K.; Harata, M.; Mizuno, S.
J. Cell Sci. 110, 3031-3041, 1997
A:Title: A nuclear matrix-associated high molecular mass nuclear antigen, HMNA, of ch
A:Reference number: 218955; MUID:9803440
A:Accession: T18535
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1151 <SHI>
A:Cross-references: EMBL:D88440; NID:d1177138; PID:d1025045; P1DN:BA424137.1

Query Match 13.4%; Score 671; DB 2; Length 1151;
Best Local Similarity 25.9%; Pred. No. 1.1e-21;
Matches 284; Conservative 100; Mismatches 419; Indels 292; Gaps 55;

```
QY 14 TPKP-----PYVDEAGSLDNGDFKVTPTDSTTOHNKYST--PKITAKPINPSS 64
D 113 TTPPQAGPAGPFPPOGAAGAPKGDGTAPGSGTNSGAGKCAADVPKATTA--ATEAP- 170
QY 65 LPNPDSTKESLSLVNKKETVETKETTTNK--QTSIDGKERKTSAKETOIEKTSADL 122
D 171 -----ASAASPTVPKATATATVTAASOGAPAAATDAAVTA--SQSAPKATV--EV 219
QY 123 APTSKVLAKPTPKAETTKGPAITTPKPTP--TPKKEASTTPKEP-----TPTTIKSAPT 177
D 220 KPAATAVAKEKATVATAAATAKATAEAKPAPVTSPTTTPCSSAEAKPLTAASPTASKA--T 277
QY 178 TPKEPAPT-----TKSAPTPKKEPAPTTPK----- 203
D 278 AEAKPVPAATASIMATKYTAEAKPAPSPSVKATTDCAVATATAPKAGPDVKAIVAVCAEA 337
QY 204 EPAPTTPKEPAPTTPKAPPTTTSAPTT-----PKEPA-----PTTPKKRAPPTTPK 250
```

Db 769 PPAPLSSPPAPQVYKSSPPVQVSSP -PPAKSSPLAVSSPPQVEKTSPPAPLSSPP 827

Qy 412 TTPKKTAP-----TTP-----KLTITTPPKLAPTTPEKAP-----TTPPELAPTTPEPT 458

Db 828 LAKSSPPHVVVSSPPVYKSSPPAPVSSPPPLPKAPSPAHVSSPPVYKSPSTP--PA 885

Qy 459 PTTPEPPATTPPKAAPMTPKKE--PAPTPPKKEPAPTPPKKEPAPTPPKKEPAPTP 517

Db 886 PTTVISP--PSEKSSPPPTPVSLPPIYKSSPPAMVSSP--PMTKSSPPVAVSSPPPT 943

Qy 518 IKE---PA-----PTTPKK---PAKELAPTTKE---PTSTSKPAPTTPKGAPTT 563

Db 944 VKSSPPAPVSSPPATPKSSPPAPVNLPPPEVYKSSPPPTPVSSPPA---PKSSPPAP 1000

Qy 564 -KEPAPTTKEPAPTTPKATPTLKEPAPTPPKKAPAPLAPTTKGTSTTSKAPPT 622

Db 1001 MSSPPEVKSPPAPVSSPPVYKSSPPAPVSSP---PPVKSPPAPVSSPPPV 1057

Qy 623 -TPKETAPTPKEPAPTPPKKAPPTTTPPTTSEVSTPTTKKEFTTHKSPDESTPL 681

Db 1058 KSPPPAPLSSPPPVYKSSPPAPVSSP--SPPPVYKSSPPAPVSSPPVYKSSPPAPV 1115

Qy 682 SAEPPTKALENSPEKGVPTTKTPATKPEMTTAKDTEEDLRTTPT 731

Db 1116 SPPAP-----VKPSLP--PPAPVSSPPVTPAPPKKEOSLPPAP 1158

RESULT 6

T30826

nascant polypeptide-associated complex alpha chain, muscle splice form gp220 - mouse

N:Alternate names: alpha-NAC protein

C:Species: Mus musculus (house mouse)

C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 24-Nov-1999

C:Accession: T30826

R:Yotou, W.Y.; St-Arnaud, R.

Genes: Dev. 10, 1763-1772, 1996

A:Title: Differential splicing in of a proline-rich exon converts alphaNAC into a muscle

A:Reference number: Z20889; MUID:96312450

A:Accession: T30826

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 12187 <YOT>

A:Cross-references: EMBL:U048363; NID:g1666688; PID:g1666689; PID:NAA18732.1

C:Genetics:

A:Gene: Naca

A:Map position: 10

A:Insertions: 24/1; 1996/1; 2024/3; 2050/3; 2099/3; 2142/3; 2183/3

A>Note: differential splicing converts alphaNAC into a tissue-specific DNA-binding activ

C:Keywords: alternative splicing; DNA binding; transcription factor

Query Match 15.5%; Score 776; DB 2; Length 2187;

Best Local Similarity 27.7%; Pred. No. 7.2e-26;

Matches 288; Conservative 122; Mismatches 389; Indels 240; Gaps 49;

Qy 34 VTTPDTSTTOHAKSTSPKITTAKINPRSLPNSDTSK-----TSLTVNKETT 84

Db 806 VQSKRVDPIMSDVTPSPKTSATVAV--PK-----DTSALSLKSVATVSLSPKAPV 857

Qy 85 VETKEETTTKOTSTDKKETTSAKETSIRKTSAKDLAPTSKVLAKPTPKAETTTGPA 144

Db 858 APSNEATVPEIPTSLKNALAAATPKETLATSIPTKVTSPS----PQTPKSVSLKCAPA 913

Qy 145 LTT-----PK-----PTPTPKKEPAPT-----PEPPTPT 171

Db 914 MTSKATEIAASKDVSPSQFPEKVPQLQHPVPTSPKSPVSDTSLGALTSPPGCP PAT 972

Qy 172 IKSAPTPKE--PAPTTKSAPTTK-----EPATTTKEPAPTTPKKEPAPTTKEP 221

Db 973 LAEPPTYKKSPPAKSAPKATATPEBEVTAAPLEIPCKSKAPATAFKSSATSSSKR 1032

Qy 222 APPT--TKSAP-----TPKEPAPTPPK--PAPTTPKKEPAPTT 257

Db 1033 APKAVSKETPSKGVTAAPLEISLPLEKTSKATPGEKSASSPKRSKTAGPK-----TPP 1089

Qy 258 KEPTPTPKKEPAPTTKEPAPTTKEPAPTTAKKAPTT--PKKEPAP-----TTPKEPAPTT 311

Db 1090 GCVIAVPEELISLPKEPTQONATPNEISLAASSOKSPKTSVPEKPEPVGVTAMPLEISAP 1149

Qy 312 TKEPSPTPK-----EPAPTTKSAPTTTKEPAPTTT--KSAPTTPK-----EPSPT 357

Db 1150 OKAPTAAPVKKOILPPEDAVTLILAGSPLSPKASKTAAKKEPAPISGVIAVSEISPS 1209

Qy 358 TKEPAPTTKEPAPTTPKKAPTT--PK-----PAPTTKEPAP-----T 396

Db 1210 KTSKTAAPKENSATLPPKRSPTAAKETPATSSEGVTAAPVSEISPPTPASKGVPT 1269

Qy 397 TTKKAPAPKKEPAPTTTKEPAPTTPKLPTPEKLAAPTTPEK--PAPTTPEELAPTTPE 455

Db 1270 LTPKAPNALAE--SPASKVKVKTAAPESTTP-----SPOKIPKAGKESAITPS 1332

Qy 456 EPTPT-----PEEPAPTTPKAAAPNTPEKAPTTPKKE--PAPTTKEPAPTTPKETA 506

Db 1323 KTPPTAAPKETSASSEGVTAAPLEIPSPKAPATAKKEPAPPS--PEGATTAAPVOIP 1380

Qy 507 PTPKGTAPTTLKEPAPTT-----KPPAKELAPTTTKEPT-- 544

Db 1381 PSPKSGSKAGSKE--TPPTPSSEGVTAAPLEIPSSKTSKMAKSEKTLVTPSSKLSOT 1439

Qy 545 -----TTSOKPAPTTPKGAPTTKEPAPTTPKKE--PAPTTPKGTAPTTLKEPAPTTPKKPA 599

Db 1440 VGPKESTLEGATAVPLEIPPSHKAKPVYDQVLTSPK--DAPTTLAE--SPSSPK-A 1496

Qy 600 PKELAPTTKGPSTSDSKAPPTKETAAPTTPKKEPAPTTPKKAPTTPEP-----PPT 654

Db 1497 PTKAAPPSR--VTYVPEKPA--TPKASGTSKVPVPEVPEVAVSSREPTVPAPV 1554

Qy 655 TSEVSTPTTK-----EPTTHKSPDESTPELSAEPKALENSPKKE--PGVPTTKTPAA 707

Db 1555 KNSSHKTSKTELEKAPATLPPSPKSPKIPSSKAPRT--SAKPEPASPSTK--PYT 1611

Qy 708 TKEMTTKDKDTERDLRTPEPTTAAPKMTKETATTEKTESKITATTOVTSTTQ 767

Db 1612 T--SLQOTAPPSLOKAPSTTIKRENAAPV-----LPVSSKAPAPARASASLSTAPAP 1665

Qy 768 DTPPKITLKTTLAPKVTTKITTTEINKPKPEKPKDRATNSKATTPKPKPK 827

Db 1666 QTPAKBATITPSCKAATETIETSTAPSLDEAPKETSE---TSSVKVLMSSP----- 1716

Qy 828 APKKPTSTKKPKMPVAVRKKTTPPKMTSTMPLELPTSRILAEALQTTTPNQTPNSK 887

Db 1717 -PKASSSKASTLP-----ATILPSLKEASVLS-----PTATSSGK 1752

Qy 888 LVEVNPKSEDAAGAGETP 906

Db 1753 DSHSPVS--DACSTGTTT 1770

RESULT 7

S48478

glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast (Saccharomyces cerevisiae)

N:Alternate names: extracellular glucanoylase; mucin-like protein MUC1; protein YIR01

C:Species: Saccharomyces cerevisiae

C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 12-Nov-1999

C:Accession: S48478; A26877; B26877; S27281; J06123

R:Rowley, K.

A:Submitted to the EMBL Data Library, October 1994

A:Reference number: S48478

A:Accession: S48478

A:Molecule type: DNA

A:Residues: 11367 <ROW>

A:Cross-references: GB:247047; EMBL:238061; NID:g603997; PID:g763364; GSPDB:GN00009;

R:Yamashita, T.; Nakamura, M.; Fukui, S.

J. Bacteriol. 169, 2142-2149, 1987

A:Title: Gene fusion is a possible mechanism underlying the evolution of STA1.

A:Reference number: A91831; MUID:87194600

RESULT 4
T16251
hypothetical protein F35A5.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 18-Feb-2000
C:Accession: T16251
R:Leinbach, D.
submitted to the EMBL Data Library, January 1996
A:Description: The sequence of C. elegans cosmid F35A5.
A:Reference number: 218485
A:Accession: T16251
A:Status: preliminary
A:Molecule type: translated from GB/EMBL/DBJ
A:Residues: 1-1274 <LEI>
A:Cross-references: EMBL:U46675; NID:g116613; PID:g116621; PIDN:AAB52641.1; GSPDB:GNOC
A:Experimental source: strain Bristol N2; clone F35A5
C:Genetics:
A:Gene: CESP:F35A5.1
A:Map position: X
A:Introns: 1272/2

Query Match 16.0%; Score 802; DB 2; Length 1274;
Best Local Similarity 28.0%; Pred. No. 3.5e-27;
Matches 290; Conservative 102; Mismatches 414; Indels 230; Gaps 56;

```
QY 6 KNRTRK-KP-----TP-----KPYVDAGSLDNGD---FKVTPPTSTTQHNVKSTS 50
DB 274 KNRTRKMWKRWEDDETPVEVEKPEPVKAPVKKKKDPAKAKANDPSPKAAKPKVPS 333
QY 51 PKITTAAPRI-NRPSLPPNSDSKETSILVNKETVEIKETTTTKQKSTGCKETTSK 109
DB 334 SPVVPPTPKNVKPKKPPWEDDEPAEVEKPKSAPEKKTIVLKAKKEPSSTPSSDPS 393
QY 110 ETQSTIEKTSANDLAPTSVLAKPPTKAE-----TTTKGPA-----LTTTP 148
DB 334 PKKAPAVKPRDSSPKKATPIQADPKAQEVPPTPVKKNPVKKYPWEVDEDDPEVEKQP 453
QY 149 KEPTPTP-----KEPASITPKPE-----TP-TIKSAPITPKPEAPTTTKSAPITPK 199
DB 454 EAPAKKTIVLKAKKEPAADTKPATSKTPEPEKKDPVKPRDSSPKKVAAPDSQAQPA- 512
QY 200 TTTKEPA-----PTTKEPAITTKKEPAPT-----TTKSAPITP 233
DB 513 TPVKNPVKMWKRPWEDDETPADVDKPTDAKKTPSLAKKDDAPAKESIKPAADTKAAR 572
QY 234 KEP-----APTTPKKRAPITPKPE-----PAPTTPKEPTTPITPKKEPAPTTPKEP 283
DB 573 RDPSEKVAAPTAPPEKKTIVLAKKKEPAGPADSKTEKEPEKSKRDPSPKKAVPAKVPKTEV 632
QY 284 APTAPKKRAP-----TTPKEPAPTTPKEPAPTTPKEP-----SPTTKEPAPTITK 333
DB 633 APAVVKKEPISKPKDITAPKKAEPNSVP-PTPVKNPVKMWKRPWEDDDAPAKVSLPE 691
QY 334 TTKPEAPITTKSAPITP-KEPSPTTKEPAPTTPK--EPAPTTPKKRAP--TTPKEPAPT 388
DB 692 PEKKTIVLAKKAPTKPDSEAADPVSGPSSKDKLAKKAPVPRDPSKAVPIKAPK 750
QY 389 TPKEPAPTTPKRAPTA-PKEPAPTTPKEPAPTTPKKTITTP-----EKLAPITPK 442
DB 751 T--EVPAAVVKKEPVAKSDPSKKK-AEPPSP--VVPPTPVKKNPVKMWKRPWEDDDA 805
QY 443 PTTPEELAPITPEEP-----TPTTPEEPAPTTPKAAAPNTPKEPAPTTPKEPAPTTP- 496
DB 806 PAEAVNVPPEBEKKTIVLAKKTPVAKRDPSPKKAAPKAPSKTDAAPSVKAPKEPVSKPE 865
QY 497 PAPTTPKETAPITPKGAPATLKEPAPTTPKKRAPKELAPIT-TTKEPTSTTSKRAPITTP 555
DB 866 PSKKKAEPPNSVP-----PTPVKNPVKMWKRPWEDDEDETEVKKPSD--PEKKTIVLA 918
QY 556 KGTAAPTTPKRAPITPKKEPAPTTPKGTAPITLKEPAPT-----TPKKRAP----- 603
DB 919 K-KEPEPKD-APKVAAPKPRDPSKKAAPVPE--KEPAKVAAPKPRDPSKKAIPANTOA 974
```

```
QY 604 APTTKGPTS-----TTSKRP-----APTTPKET-----APTTPKEPAPTTPKK 646
DB 975 PPTPVKNPVKMWKRPWEDDEPAEPVSAPEBEKKTIVLAKKAPKAPKRPD-----SRK 1031
QY 647 TPETPPTTSVSPTTTPKKEPTTIHKSP-----DESTPLSA-EP--TPKALENSPK 698
DB 1032 AAK-PDPAIPEV-PPTPVKNPVKMWKRPWEDDESEPSASAPBEKKTIVLAKKAP 1089
QY 699 V-----PTTKPAAT-----KPEMTITAKDKTEEDLRTPTPT--TAPK 737
DB 1090 TKPDSEAADPVSGPTSKDPLSKKAPYEKPTTDDKDLKPSPAKPEKAPAPK 1149
QY 738 MKETATITTEKTSKITATTQVSTTTQDTPEFKITITTLTKTILAPKVTITTKTITTE 797
DB 1150 KWKPVWDDPDPEPADFTVPAPSKRPDDEPADPLG-----GPKTKDPK----- 1193
QY 798 IMNKPEETAPKDRATNSKATTPKPKPKPTSTKPKTPRVKPKK----- 848
DB 1194 -LNKKAPAEKPTK-----PKKEVSKPEPKPTPEPKR-APKMKRPWEDDDPEDE 1243
QY 849 ---TTPPRKMTSTMP 861
DB 1244 ADFTMAPAKKPDTEDP 1259
```

RESULT 5
S49915
extensin-like protein - maize
C:Species: Zea mays (maize)
C:Date: 05-Mar-1995 #sequence_revision 12-May-1995 #text_change 29-Oct-1999
R:Rubinstein, A.L.; Broadwater, A.H.; Lowrey, K.; Bedinger, P.A.
submitted to the EMBL Data Library, June 1994
A:Description: Pex genes: pollen-specific genes with extensin-like domains.
A:Reference number: S49915
A:Accession: S49915
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1188 <RUB>
A:Cross-references: EMBL:Z34465; NID:g600117; PIDN:CAA84230.1; PID:g600118

Query Match 15.6%; Score 784; DB 2; Length 1188;
Best Local Similarity 29.9%; Pred. No. 1.9e-26;
Matches 230; Conservative 59; Mismatches 359; Indels 122; Gaps 33;

```
QY 13 PTPKPYVDAGSLDNGDPKVTTPPTSTTQHNVKSTSPKITTAAPRINRPSL----- 68
DB 460 PTPHSPPAD-----DYPPTPVKPKSPPATSPSPQVPAPASPPPSLVLSPEQ 510
QY 69 SDTSKETSILVNKETVEIKETTTTKQKSTGCKETTSKAKETQSTIEKTSAKDLAPTSKV 128
DB 511 APVG-----SPPPVKTTSPPAIG-----SPSPPPVSVV 541
QY 129 -----LAKTPKAEITTKGPAITTPKEPTTPKKEPASTTPKEPTTIKSAPTTPKEPA 183
DB 542 SPPPVKSPPPAPVSGSPPPKSPPPAPVASPPPVKSP--PPPTIVASPPPVKSP 599
QY 184 PTTKSAPITP--KEPAPTTPKEPAPTTPKKEPAPTTPKKEPAPTTPKSAPTTP-KEPAPT 240
DB 600 PPAVASPPPVKSPPPPTPVASPPPAVASPPPMKSPPPPTVSSPPPEKSPPPPP 659
QY 241 PKKAPITTPKEPAPT-----PKC-----PTPTTPKEPAPTTPKEPAPTTPKEPAPTAP 291
DB 660 PAKSTPPEEYPTPTSVKSSPPPEKSLPPPTLISPPPOKPTPESTPSAP-PSSEPK 718
QY 292 APTTPKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKSAPTTPKKEPAPTTPK 351
DB 719 SP--KPEVSSPQTP-----KSSPPAVSSPPPTPVSSPALAPVSSPVKSSP----- 768
QY 352 EPSPTTPKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPT 411
```


Db 865 SDEPTSPDEPTSP--DEPTSP--DEPTP--SDEPTSPDEPTSPDEPTSPDEPTSP 919
QY 302 TTPKPEPA--PTTKPESP--TKPEPAPTTKS--APTTKPEPAPTTKSAPTTKPESPPT 357
Db 920 SDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSP 968
QY 358 TKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPA 416
Db 969 SDEPTSP--DEPTSP--DEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSP 1019
QY 417 TAPTKPKLPTTKPEKAPPTTKPEKAPPTTKPEKAPPTTKPEKAPPTTKPEKAP 475
Db 1020 -----EPTSPDE--PTSPDEPTSPD--EPTSPDEPTSPDEPTSPDEPTSP 1064
QY 476 NTPEKAPPTTKPEPA--PTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPE 532
Db 1065 TPSEDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSP 1122
QY 533 ELAPTTKPEPT--STSPDKAPPTTKGAPPTTKPEPAPTTKPEPAPTTKGAPTTKPE 591
Db 1123 D--EPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSP 1175
QY 592 PT--TPKPAKELAPTTTKAPTSTSKAPAP--TPKELAPTTKPEPAPT--TPKPAPT 646
Db 1176 PSETPPEP-----IPTDPSDEPTSPDEPTSPDEPTSPD--EPTSPDEPTSPDEPTSP 1229
QY 647 TPPTPTPTSEVSTPTTKPEPTTKPEPTTKPEPTTKPEPTTKPEPTTKPEPTTKPE 706
Db 1230 DTPEDEPTSPD--EPTSPDEPTSPD-----PSDEPTP--SETPPEP--IPTDPSD 1277
QY 707 ATKPEMTTAKDKTTERDLTTPPTTAAKPKMKETATTEKTESKITATTOVSTTI 766
Db 1278 EPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSP 1337
QY 767 QDTPPTPTTKTTTTLAPKVTTKTTTTEIMKPEETAKPKDRATNSKATTPKPO--KP 825
Db 1338 DEPTSPDEPT-----PSDEPTSPDEPTSPDEPTSPDEPTSP 1372
QY 826 TKAKRKPTSTKKPKTPMPVRKPKTTPPKMTS-----TMELNPTSKIA 870
Db 1373 SEPPEDEPTPTTP-----TPPTSTPTSPSGSGSGSGSGGAGTVPSTPTPTSP 1424
QY 871 EAMLOTTTRPNOTRPSKLYEVNPKSEDAAGETPHMLLRPHYEMPEVTDMOYL 927
Db 1425 -----KPTSTAPPEIE--EPTSPDVGAIGGHAHYLKG-----PDGSRP 1465

RESULT 3
T31108
cyst germination specific acidic repeat protein precursor - Phytophthora infestans
C:Species: Phytophthora infestans (potato late blight agent)
C:Date: 22-Oct-1999 #sequence:rev150n 22-Oct-1999 #text:change 22-Oct-1999
C:Accession: T31108
R:Goernhardt, B.
Submitted to the EMBL Data Library, April 1998
A:Reference number: Z20986
A:Accession: T31108
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1489 <GOE>
A:Cross-references: EMBL:AF061185; NID:g3851513; PID:g3851514; PIDN:AA072308.1
C:Genetics:
A:Gene: car90

Query Match 17.0%; Score 853; DB 2; Length 1489;
Best Local Similarity 32.3%; Pred. No. 2.9e-29;
Matches 355; Conservative 46; Mismatches 475; Indels 222; Gaps 49;

QY 9 TTKKPTPKPPVVDGSGDNGDFKVTTPDTSTT-----QHNKYSTSKITAKINR 62
Db 338 TPAPTEKEVDYETTYVTEESTYAPTCKSETNAPTERMHYAHLEKPCDTEVTWYAPTEET 397

QY 63 PSLP-----PNSDT-----SKETSLVNKETTVETKETT-----TNKOTSTD 100
Db 398 TPAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAP 457
QY 101 GKEKTTSAKETOSIKTSADKLAPTSKYLAKPEKAEKTTKQALTYKPEPTTKPEPA 160
Db 458 TEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAP 514
QY 161 STTPKEPTPTTKKSAPTTKPEPAPTTKSAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPE 217
Db 515 EETPYEPTTEET--TYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYA 573
QY 218 TKPEPAPTTKSAPT-----TPKPEPAPTTKKAPPAPTTKPEPAPTTKPEPTTKPE 267
Db 574 TYAPTEET--YAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYA 632
QY 268 P-----AP-----TKPEPAPTTKPE-----APAPKPKAPATTKPEPAPTTKPE 304
Db 633 PTEETTYASTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYA 692
QY 305 KEPAPT--TTKESPTTKPEP-----APT--TKKSAPT-----TKPE-----A 339
Db 693 TTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYA 752
QY 340 PT--TKKSAPT--TPKESPTTKPEP-----APTPEKPAPTTKKAPPAPTTKPEPAPT-- 388
Db 753 PTEATTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 812
QY 389 TPKEPAPTTKKAPPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAK 448
Db 813 TPPEPTTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYA 870
QY 449 LAPPT--TPPEPTPTTKPEP-----APT-----TPKAAPNTPEKAPPAPTTKPEP 489
Db 871 YAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYA 930
QY 490 -----ADTPKEPAPTTKKAPPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAK 539
Db 931 EETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYA 990
QY 540 KEPTSTSDKAPPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAK 585
Db 991 YAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYA 1050
QY 586 -----TLKPEPAPTTKPKP-----APKE-----LAPPT--TTKGTSTTSKAPPAPT 624
Db 1051 EETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYA 1110
QY 625 KETAPPTTKPEPAPTTKKAPPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAK 682
Db 1111 YAPAEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYA 1169
QY 683 AEPKPA--LENSPKPEP-----VPTTKPAKPEMTTAKDKTTERDLKTP-----ETT 732
Db 1170 TEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYA 1229
QY 733 TAAPKMKETATTEKTESKITATTOVSTTODTT-----PKKIT--LKT 779
Db 1230 TYAP-----TEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYA 1284
QY 780 TTKAPKVTTKTKITTTTELIMKPEETAKPKDRAT--NSKATTPK--OKTPAKPKPT-- 833
Db 1285 TTYAPTEATTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYA 1344
QY 834 -----STKPK--KTPMPVRKPKTTPPKMTSMBELNPTSKIAEAMLOTTTRP--N 881
Db 1345 PAEBSTSTVSTKPCNTLEPTDEPTDDE--PSDEPTDEPTDPLPDEBSPDCDN 1402
QY 882 QTPNSKLYEVNPKSDDAG 899
Db 1403 OGINGIGVENKVRVNNAG 1420

J. Clin. Invest. 87, 77-82, 1991
 A:Title: Human bronchus and intestine express the same mucin gene.
 A:Reference number: A61257, MUID:91086481
 A:Accession: A61257
 A:Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1,1925-1948, 'TTS', 1952-1954 <JAN>
 A:Experimental source: bronchus
 A:Biochem. Biophys. Res. Commun. 183, 821-828, 1992
 A:Title: Human intestinal mucin-like protein (MLP) is homologous with rat MLP in the C-terminus
 A:Reference number: P00328, MUID:92198477
 A:Accession: P00328
 A:Molecule type: mRNA
 A:Residues: 2328-2468 <XUG>
 A:Cross-references: GB:M85523
 A:Experimental source: small intestine
 A:Accession: P00329
 A:Molecule type: protein
 A:Residues: 2328-2342, 'K', 2344-2354 <XUG1>
 A:Gene: GDB:MUC2
 A:Cross-references: GDB:120203; OMIM:158370
 A:Map position: 11p15.5-11p15.5
 C:Superfamily: von Willebrand factor; von Willebrand factor type A repeat homology; von F.2766-2834/Domain: von Willebrand factor type C repeat homology <VWC>

Query Match 21.2%; Score 1062.5; DB 2; Length 3020;
 Best Local Similarity 29.2%; Pred. No. 8e-38;
 Matches 342; Conservative 74; Mismatches 442; Indels 315; Gaps 39;

```

QY 32 EVATTPDSTTQH---NKVSTSPKTTAKPINRPSLPSNDSKTSLSLVNKKETVETK 88
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 1240 WEICPNCNTEVHEKFNICSTTRPSLTITFTTITLPTTP-----TSFTTTTTTTTTPTS 1291
QY 89 ET--TTTNK-----QTSNDGKEKITS--AKETOSIEKTSAND----- 121
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 1292 STVLSTTKLCLCMSDWINEHPSGSDGDREPDGVCAGAEIECNVNDPRLSLEQH 1351
QY 122 -----LAPTSKVLAKPTPKAE 137
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 1352 GOKVOCDSVGFICKNEDQNGFGLCYDYKIRVNCQWMDKCTITTPSPPTTPSPPT 1411
QY 138 TTTGPAITTEKEPPTTPPKPASTTKEPPTTITKSAP--TTKEPAPTTTSAPPTTKE 196
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 1412 TTTTLEPPTTSPPTTTPPTTTPPTTTPSPPTTTPPTTTPPLPTTTPSPPISTTTTTPPTT 1471
QY 197 PAPPTTKEPAPTTKEPAPTTKEPAPTTTSAP--TPKKEPAPTTTPPKPAPTTKEPA 255
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 1472 P-PTTTPSPPTTSPPT--TTTTPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTP 1529
QY 256 TPKEPTTPTEKEPAPTTKEPAPTTTPKPAAPKPAPTTTPKEPAPTTTPKEPA 315
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 1530 TTPPEPTTTPSPPTTTPITP--PTSTTTLPLPTTTPSPPTTTPPTTTPPTTTPSP 1588
QY 316 SPPTTKEPAPTTTSAP--TTTKEPAPTTTSAP--TPKESPTTTPKEPAPTTTPKE 373
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 1589 PTITTTTPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTP 1648
QY 374 PKKPAPTTKEPAPTTTPKEPAPTTTPKPAAPKPAPTTTPKEPAPTTTPKEPA 433
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 1649 TTPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTP 1700
QY 434 APPTTKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKPAAPKPAPTTTPKEPA 493
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 1701 PTITTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTP 1759
QY 494 PKKEPAPTTTP-----KETAPTP-----KG----- 512
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 1760 PLPPSTTPPTTTPPTTTPPTTTPCVPCLCNMTGMLDSGKPRFHKRGDTLIGDVCGRGMA 1819

```

```

QY 513 ----- 512
Db 1820 NISCRATWPDVPIQLGTVCDVSVGLICKNEDQKRGVITPAFCLNFINOCCECV 1879
QY 513 TAPPTLKEPAPTTTPKPAKELAPPT--TKEPSTTSNDK---APT--TPKGAPTTTPKE 565
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 1880 TQPTTW-----TTTTENPPTPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPT 1936
QY 566 PAPPTTKEPAPTTTPKGAPTTTPKEPAPTTTPKPAKELAPPTTTPGPTST--TSDKPAPTTP 624
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 1937 PTPGTQTPPT--TTPLTTTTTTPPTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPT 1995
QY 625 KETAPTPKEPAPTTTPKPP-----APTPPTPTTSEVSTPTTPK-----EPTTHK 672
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 1996 ITTTTTPVATPTPTGTQTPPTMIPISITTTTPVPTPTTTPGTPPTTHSTAPAEITSNP 2055
QY 673 SPDESTPLSAEPTPKALENSPKRPGVPTTTPATPTPEMTTAKDKTTERDLRTTPPT 732
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 2056 PPESTTPQTSSTSPLESTTLSTLPALEMTSTAPSTTPAPPTTSGHTLSPPST 2115
QY 733 TAPAKMTKETATTEKTESKITATTTQVSTTTQDTPPKITT---LKTTLAPKVTTP 789
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 2116 TTSPTGTPTRGTTP--GSSAPPTSTVQTTTTSAMPTTTPPLSTPILRTTGLRPTPSV 2173
QY 790 KKTITTEINMKPPE-----TAKPKDRATNSKATTP 820
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 2174 LICCVLDNYTAAEGEYVNGTYGDTCYFVNCSLSTLEFYNMCSPTSPPTPPSK--STP 2232
QY 821 KPOKPTAPKPKPTSTKPKTMPRVRKPKTPPTPKKMTSTPELNPSTRLAEMLOTTPR 880
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 2233 TPKSPSTSPKPPPGKTPPCPDPPR-----QENETWLCDCRW--ATCKR 2278
QY 881 NQPNKSLVEVP-----KSEDAAG-----AEGETPHML----- 910
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 2279 NMTVEIKVACEPPMPFCSNGLOPVRYVEDPDCQWHECDXCCTGMGD--PHVVTDPGLX 2337
QY 911 -----RPHVMEPTVPMQDYLPRVPMOGLIN 937
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 2338 YSYQNGNTIYVLEISPSVD-----NFGVYID 2364
```

RESULT 2
 T18262
 S-layer protein - Clostridium thermocellum
 C:Species: Clostridium thermocellum
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T18262
 R:Juhn, T.; Beguin, P.; Aubert, J.P.
 J. Bacteriol. 175, 1891-1899, 1993
 A:Title: Organization of a Clostridium thermocellum gene cluster encoding the cellulose.
 A:Reference number: 218847; MUID:93209931
 A:Accession: T18262
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1664 <FGU>
 A:Cross-references: EMBL,X67506; NID:g296879; PID:g296881; PIDN:CAA47841.1

Query Match 18.8%; Score 942; DB 2; Length 1664;
 Best Local Similarity 33.0%; Pred. No. 5.6e-33;
 Matches 276; Conservative 83; Mismatches 312; Indels 166; Gaps 43;

```

QY 128 VLAKPTP--KAETTTGPAITTEKEPPTTTPKEPASTTKEPPTTITKSAPPTTKEPA 186
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 758 VVIOAPAIKASDEPIPDITSDPTSP-----DEPTPS---DEPTSDPTSP 804
QY 167 TKSAPTTPKEPAPTTTPKEPA-----PTTTPKEPAPTTTSAPPTTTPKEPAPTTTPK 242
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 805 EPTPSETPEEPIDPDITSDPTSDPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPT 864
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 243 KPAPTTPKEPAPTTTPKEPTTPTPKEPAPTTTPKEPAPT--TPKPAATKAPKPAPTTPKEPA 301
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:13:30 ; Search time 82.85 seconds
(without alignments)
865.181 Million cell updates/sec

Title: US-09-556-246-1_COPY_200_1140
Perfect score: 5011
Sequence: 1 VKDNKKNRKTKKPPV.....DMDYLPVYVNOGIINPMLS 941

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 segs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR.*
2: PIR.*
3: PIR.*
4: PIR.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	1062.5	21.2	3020	A43932	mucin 2 precursor, S-layer protein -
2	942	18.8	1664	T18262	Cyst germination s
3	853	17.0	1489	T31108	hypothetical prote
4	802	16.0	1274	T16251	extensin-like prot
5	784	15.6	1188	S49915	nascant polypeptid
6	776	15.5	2187	T30826	glucan 1,4-alpha-g
7	762.5	15.2	1367	S48478	high molecular mas
8	671	13.4	1151	T18535	mucin 1 precursor,
9	666.5	13.3	1344	A5175	hypothetical prote
10	659.5	13.2	1229	T25697	extensin class 1 p
11	633	12.6	489	T11622	hypothetical prote
12	632	12.6	761	C84672	hypothetical prote
13	632	12.6	7962	T38346	elastic titin - hu
14	629	12.6	3507	T34511	hypothetical prote
15	626.5	12.5	990	T15618	nucleolar phospho
16	622.5	12.4	839	T75518	hypothetical prote
17	622.5	12.4	6642	F75518	protein UNC-69 - C
18	620.5	12.4	971	T19431	hypothetical prote
19	617	12.3	3570	T45025	mucin MUC5B, trach
20	607.5	12.1	801	T29018	gene LF3 protein -
21	607.5	12.1	924	S27923	larval glue protei
22	605	12.1	379	S50125	hydroxyproline-ric
23	559.5	11.2	350	S22456	hypothetical prote
24	556.5	11.1	856	T16543	polyphenolic adhes
25	555.5	11.1	875	S23760	ascites stialoglyco
26	551	11.0	1630	A55577	hydroxyproline-ric
27	549	11.0	620	S06733	hypothetical prote
28	549	11.0	2232	T34434	calphostin - fruit
29	545	10.9	873	A47283	

30	542	10.8	369	2	S20500	hydroxyproline-ric
31	538.5	10.7	416	2	J00465	extensin precursor
32	533	10.6	1087	1	QFMSH	neurofilament trip
33	532	10.6	756	2	T27642	hypothetical prote
34	530.5	10.6	1162	2	JH0557	exo-alpha-stalidas
35	530	10.6	865	2	A47282	calcium-binding pr
36	522	10.4	328	2	J00985	hydroxyproline-ric
37	518	10.3	813	2	S70795	USA protein precu
38	518	10.3	1072	1	A37221	neurofilament trip
39	512	10.2	866	2	T45462	membrane glycoprot
40	509.5	10.2	662	2	A45155	mucin F1W-C.1 - Af
41	509.5	10.2	867	2	T45463	membrane glycoprot
42	509	10.2	1459	2	T32271	hypothetical prote
43	505	10.1	700	2	A54641	interspersed repea
44	504.5	10.1	606	2	A43427	neurofilament trip
45	503.5	10.0	1118	2	A48292	mucin, tracheobron

ALIGNMENTS

RESULT 1

A43932
mucin 2 precursor, intestinal - human (fragments)

N:Alternate names: mucin SMUC-41
C:Species: Homo sapiens (man)
C>Date: 10-Mar-1993 #sequence, revision 12-Apr-1996 #text, change 05-Nov-1999
C:Accession: A49963; A45106; B45106; A43932; B33532; A61257; P00328; P00329
R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, B.; Kim, Y.S.
J. Biol. Chem. 269, 2440-2446, 1994
A:Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of t
A:Reference number: A49963; MOID:94132002
A:Accession: A49963
A:Molecule type: mRNA
A:Residues: 1-639 <GUT>
A:Cross-references: GB:M412198
R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Rothe, E.M.; Lagace, R.E.; Kim, Y.S.
J. Biol. Chem. 267, 21375-21383, 1992
A:Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both up
A:Reference number: A45106; MOID:93016075
A:Accession: A45106
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 2037-3020 <GUT>
A:Cross-references: GB:M94131; NID:g186395; PIDN:AAA59163.1; PID:g186396
A:Note: sequence extracted from NCBI backbone (NCBIP:116698)
A:Experimental source: colon
A:Note: sequence extracted from NCBI backbone (NCBIP:116698)
R:Toribara, N.W.; Gum Jr., J.R.; Culhane, P.J.; Lagace, R.E.; Hicks, J.W.; Petersen,
J. Clin. Invest. 88, 1005-1013, 1991
A:Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polym
A:Reference number: A43932; MOID:91358717
A:Accession: A43932
A:Molecule type: DNA
A:Residues: 1343-1350, 'L', 1352-1411, 'S', 1413-1448, 'P', 1450-1503, 'T', 1505-1915 <TOR>
A:Cross-references: GB:M74027; NID:g188863; PIDN:AAA59875.1; PID:g188864
A:Note: sequence inconsistent with the nucleotide translation
A:Note: sequence extracted from NCBI backbone (NCBIP:55749)
R:Gum, J.R.; Byrd, J.C.; Hicks, J.W.; Toribara, N.W.; Lamport, D.T.A.; Kim, Y.S.
J. Biol. Chem. 264, 6480-6487, 1989
A:Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and eva
A:Reference number: A33532; MOID:89197956
A:Accession: B33532
A:Molecule type: mRNA
A:Residues: 1916-2193 <GUT>
A:Cross-references: GB:M22405; NID:g188873; PIDN:AAA36334.1; PID:g188874
A:Experimental source: intestine
R:Jany, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Kim, Y.S.; Basbaum, C.B.

```

DB 472 APAAHGVTSAPDXRPGSTAPAAHGVTSAPDXRP--XPGSTAPAAHGVTSAPDXRPGXS 530
QY 589 TTP-----KKPAKELAPTTTKEPTSTSDKPA--TTPKGAPTTTKEPAATTP 636
DB 531 TAPAAHGVTSAPDXRPGSTAP--XAHGVTSAPDXRPGSTAPXAHGVTS---APDXR 585
QY 637 KEPAATTPKG---TAPTTTKEPAATTP-----KKPAKELAPTTTKEPTSTSD 682
DB 586 PXPGSTAPAAHGVTSAPDXRPGSTAPAAHGVTSAPDXRPGSTAP--XAHGVTSAPDX 644
QY 683 KPA--TTPK---ETAPTTTKEPAATTP-----KKPAATTPETPTTSEVSTP 726
DB 645 RPYPGSTAPAAHGVTSAPDXRPGSTAPAAHGVTSAPDXRPGXS--TAPXAHGVTSAP 702
QY 727 TTTKEPT-----IHKSDESTPELSAEPITKALENSKEGCVPTTKPAA----- 772
DB 703 DXRPGSTAPAAHGVTSAPDXRPGSTAPAAHGVTSAPDXRPGSTAPXAHGVTSAP 762
QY 773 -TPEMTTAKD---TTERDLRTTPETTAAPPKMTKEATTTTEKTESKITATTQTVS 828
DB 763 DXRPGSTAPAAHGVTSAPDXRPGST--APXAHGVTSAPDXRPGSTAPXAHGVTS 820
QY 829 TTTQDTTPEKITTLKTTTAPKVTTTKITTTTEINMKPEETAAPKDRATNSKATTPKQ 888
DB 821 APDXRPG---STAPXAHGVTS---SAPDXRPGSTAPXAHGVTSAPDXRPGX- 868
QY 889 KPTKAPKPTSTKKPTMPVVRKPTTPTPRKMTSTPELNPSTRIAEAMLQTTTRENQ 948
DB 869 -GSTAPAAHGVTSAPDXR--XPGSTAPAAHGVTS--PDXRPX-----PGST 911
QY 949 -PNSKLEVNPKSEDAGAGETPHMLRPHVMEVTPDMYDLPRVNOGIIINPMLSD 1007
DB 912 APXAHGVTSAPDXRPGSTAPXAH-----GVTSAPDXRPGXGSA---PXAHG 957
QY 1008 ETNICKGKPRVDGLT--LRNGTLVAFRGHTYWMLSPPSPSPARITTEVWGIPSPID 1062
DB 958 VTSAPDXRPGSTAPXAHGVTS-----PDXRPGSTAPXAHGVTSAPD 1003

RESULT 15
US-08-276-967-2
; Sequence 2, Application US/08276967
; Patent No. 5851817
; GENERAL INFORMATION:
; APPLICANT: Hardy, Daniel M.
; APPLICANT: Garbers, David L.
; TITLE OF INVENTION: Species-Specific Egg-Binding Proteins of
; TITLE OF INVENTION: Sperm
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P. O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/276 967
; FILING DATE: Submitted Herewith
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Kitchell, Barbara S.
; REGISTRATION NUMBER: 33,928
; REFERENCE/DOCKET NUMBER: UTSD:418\KIT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713-787-1400
; TELEFAX: 713-789-2679

```

```

; TELEX: 79-0924
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2476 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-276-967-2

```

```

Query Match 6.2%; Score 419.5; DB 2; Length 2476;
Best Local Similarity 32.5%; Pred. No. 1.4e-19;
Matches 163; Conservative 43; Mismatches 177; Indels 119; Gaps 31;

```

```

QY 471 KPEAPTTTKEPAATTTKLLPTTPEKLAATTTPEKPAATTPPEELAPTTPEPTTPEEP 530
DB 313 PSESTSTKRPVAPTE---KPTVPSSEITYTTPKEKPMVHMEKPIVHT--EKPTVPT-EKP 365
QY 531 APTTPKAAAPNTPEKPA-----PTTPKPEAPTTTPEKPAATTPPKETAATTPKGTATTLKEP 586
DB 366 TTPTEKSTVPT--KKPTVEKEPILPPE-GPTVPAE-RPTTPREGPAVPPKG--PTVLTLE- 418
QY 587 APTTPKKAPKELAPTTTKEPTSTSDKPAATTPKGTAP--TPEKPAATTPKPEAPTTPK 645
DB 419 -----WPTSHTKSTVHTEKPIIPGKSTIPEKPMVPTKRP-----TTP- 458
QY 646 GTAPTTTKEPAATTPKPAKELAP--TTKGPSTT---SDKPAATTPKETAATTTKEPA 701
DB 459 -TEKTTTPAKRPV--PIEKPMVPTERTTPTERTTTPKEKPIVPEKLTVPPT--EKPT 512
QY 702 PTPPKKPAATTPETTPP-----TTSEVSTPT-----TTKEPTTIHKSDESTPELSAEP 751
DB 513 VPTEKPIVPTKEKTIPEKLTVLERTTPTTEKTTIPTKEKPIVPEKPSPT-EKPTVPT 571
QY 752 PKALENSKPEPGVPTTKPAATKPEMTTAKDKTTERDLRTTPEPTTAAPKMTKETATTT 811
DB 572 -----EEPTTPTKEKLTVPPT--ERTTTPTKRRTTPTIIRTTPTTI-----RTTTP 613
QY 812 EKTTEKITTATTQVSTTQDTTPEKITTLKTTTTLAPKTTTTKTTTTEINMKPEETA 871
DB 614 ERTT-----TPTRTTPTTERTT--TPKKTTP-----VPTEKTIIPPT-----ERT 652
QY 872 KPKDRATNSKATTPKPOKPTPKAPKPTSTKKPTMPVRKPTTPTPRKMTSTMPELNP 931
DB 653 AP-----TTPQP-SPTLVPTQPAVVPST-----SATTVPRTTIASC-----PN 693
QY 932 SRIAEMLQTTTRPNQTPNSKL 953
DB 694 AHFERCACPVSQ-SPTPNCEL 714

```

```

Search completed: April 26, 2002, 16:25:49
Job time: 489 sec

```

QY 773 -TKPKMTTAKDK---TTERDLRTTPEPTTAAPKMTKETATTTERTTESKITATTTQVTS 828
 Db 763 DKRPXPGSTAPXAHGVTSAAPDXRPXGST--APXAHGVTSAAPDXRPXGSTAPXAHGVT 820
 QY 829 TTTQOTTPPKITLTTLTTLAPVNTTKKITTTEIMNKPPEEFAKRRATNSKATTPKQ 888
 Db 821 APDXRPXG---STAPXAHGVT-----SAPDXRPXGSTAPXAHGVTSAAPDXRPX 868
 QY 889 KPTKAPKPTSTKPKTPMRVARKPTTPTPRKMTSTMPELNPTSLAEMLOTTTRPQT 948
 Db 869 -GSTAPXAHGVTSAAPDXRP--XPGSTAPXAHGVTSA-PDXRPX-----PGST 911
 QY 949 -PNSLVEVNPKESEDGAGAGETPPMLLRPHVEMPEVTPDMOYLPRVNOGIINPMLSD 1007
 Db 912 APXAHGVTSAAPDXRPXGSTAPXAH-----GYTSAPDXRPXGSTA---PXAHG 957
 QY 1008 ETNINNGKPVQDLT--LANGTLVAFRGHYFMWLSFSPSPARRKTTTEWGIDPSPD 1062
 Db 958 VTSAPDXRPXGSTAPXAHGVTSA-----PDXRPXGSTAPXAHGVTSAAPD 1003

RESULT 14

US-09-083-116-2
 : Sequence 2, Application us/09083116
 : Patent No. 6203795

GENERAL INFORMATION:

APPLICANT: CHAMON, Pierre
 APPLICANT: KIENV, Marie-Paule
 APPLICANT: LATHE, Richard
 APPLICANT: HAREUVENI, Marc
 TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
 TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
 NUMBER OF SEQUENCES: 5
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
 STREET: P.O. Box 1404
 CITY: Alexandria
 STATE: Virginia
 COUNTRY: United States
 ZIP: 22313-1404

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/083,116
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/479,537
 FILING DATE:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: MO PCT/FR91/00835
 FILING DATE: 23-OCT-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/039,320
 FILING DATE: 04-APR-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/403,576
 FILING DATE: 14-MAR-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Teskin, Robin L.
 REGISTRATION NUMBER: 35,030
 REFERENCE/DOCKET NUMBER: 017753-025
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 836-6620
 TELEFAX: (703) 836-2021
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2035 amino acids
 TYPE: amino acid

QY 140 TSLVNNKTEVYKETTITTKQVTSCKEETTSKETSIAKD-LAPTSTVLAKPT 198
 Db 16 TLTIV-----VSGSHASTPGCEKETSTQSSVSPSTEKNNAVNTSSLSHS 65
 QY 199 P-KAETTKGP--ALTTPKRP-----PTTPKEASTTPKKEPTPTTISAP 241
 Db 66 PGSSSTTGQODVTLAATPEPASGSANTGQODVTSVPTRPALGSTTPPAHDVT--SAP 122
 QY 242 TTPKEPAPTT-----KSAPTPKRP-----APTTPKEPAPTTTPKREPAPT 282
 Db 123 --DNKRAPGSTAPXAHGVTSAAPDXRPXGSTAPXAHGVTSAAPDXRPXGSTAPXAHGVT 180
 QY 283 TKE--PAPTT-----KSAPTPKREPAPTTPK--KRAPTPKREPAPTTPKKEPTPT 329
 Db 181 APDXRPXGSTAPXAHGVTSAAPDXRPXGSTAPXAHGVTSAAPDXRPXGSTAPXAHGVT 240
 QY 330 PKE--PAPTKRP-----APTTPKEPAPTAPK--KRAPTPKREPAPTTPKKEPAPT 376
 Db 241 APDXRPXGSTAPXAHGVTSAAPDXRPXGSTAPXAHGVTSAAPDXRPXGSTAPXAHGVT 300
 QY 377 TKEPSTTPKREPAPT--TKSAPTTTPKREPAPT--TKSAPTTTPKREPAPTTPK--E 425
 Db 301 APDXRPXGSTAPXAHGVTSAAPDXRPXGSTAPXAHGVTSAAPDXRPXGSTAPXAHGVT 359
 QY 426 PAPTTTPKREPAPTTPK--KRAPTTTPKREPAPTTPK--EAPTTTPKREPAPTTPKKEPAPT 477
 Db 360 SAPDXRPXGSTAPXAHGVTSAAPDXRPXGSTAPXAHGVTSAAPDXRPXGSTAPXAHGVT 419
 QY 478 TPKETAP-----TTPKRLTTPTEKLAPTTPKEKAPPTPELIAPTTPEEPPTTPKEPAPT 533
 Db 420 SAPDXRPXGSTAPXAHGVTSAAPDXRPXGSTAPXAHGVTSAAPDXRPXGSTAPXAHGVT 471
 QY 534 TPKA-----AAPNTPKREPAPTTPKREPAPTTPKREPAPTTPKGP--APTTLKEPAP 588

STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 FEATURE:

NAME/KEY: Peptide
 LOCATION: 128..1899
 OTHER INFORMATION: /note= "The amino acids spanning
 128 to 1899 constitute a repeated region wherein the repea
 OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such
 OTHER INFORMATION: repeats varies from 1 to 40."

FEATURE:
 NAME/KEY: Peptide
 LOCATION: 134
 OTHER INFORMATION: /note= "Amino acid 134 is X1 - Xaa
 Xaa Xaa which is the codon for Pro or Ala wherein Pro - CC
 OTHER INFORMATION: CCC, CCA, or CCG; and Ala - GCT, GCC, GCA, or GCG."
 OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr - ACT, ACC,
 OTHER INFORMATION: or ACG; and Asn - AAT or AAC."

FEATURE:
 NAME/KEY: Peptide
 LOCATION: 147
 OTHER INFORMATION: /note= "Amino acid 147 is X2 - Xaa
 which is the codon for Pro or Ala wherein Pro - CCT, CCC,
 OTHER INFORMATION: or CCG; and Ala - GCT, GCC, GCA, or GCG."

FEATURE:
 NAME/KEY: Peptide
 LOCATION: 1..21
 OTHER INFORMATION: /note= "Amino acids 1 to 21 are a -
 OTHER INFORMATION: 21 amino acid precursor sequence."

US-09-083-116-2

Query Match 6.2%; Score 424.5; DB 4; Length 2035;
 Best Local Similarity 23.7%; Pred. No. 5.3e-20;
 Matches 255; Conservative 69; Mismatches 511; Indels 241; Gaps 53;

QY 140 TSLVNNKTEVYKETTITTKQVTSCKEETTSKETSIAKD-LAPTSTVLAKPT 198
 Db 16 TLTIV-----VSGSHASTPGCEKETSTQSSVSPSTEKNNAVNTSSLSHS 65
 QY 199 P-KAETTKGP--ALTTPKRP-----PTTPKEASTTPKKEPTPTTISAP 241
 Db 66 PGSSSTTGQODVTLAATPEPASGSANTGQODVTSVPTRPALGSTTPPAHDVT--SAP 122
 QY 242 TTPKEPAPTT-----KSAPTPKRP-----APTTPKEPAPTTTPKREPAPT 282
 Db 123 --DNKRAPGSTAPXAHGVTSAAPDXRPXGSTAPXAHGVTSAAPDXRPXGSTAPXAHGVT 180
 QY 283 TKE--PAPTT-----KSAPTPKREPAPTTPK--KRAPTPKREPAPTTPKKEPTPT 329
 Db 181 APDXRPXGSTAPXAHGVTSAAPDXRPXGSTAPXAHGVTSAAPDXRPXGSTAPXAHGVT 240
 QY 330 PKE--PAPTKRP-----APTTPKEPAPTAPK--KRAPTPKREPAPTTPKKEPAPT 376
 Db 241 APDXRPXGSTAPXAHGVTSAAPDXRPXGSTAPXAHGVTSAAPDXRPXGSTAPXAHGVT 300
 QY 377 TKEPSTTPKREPAPT--TKSAPTTTPKREPAPT--TKSAPTTTPKREPAPTTPK--E 425
 Db 301 APDXRPXGSTAPXAHGVTSAAPDXRPXGSTAPXAHGVTSAAPDXRPXGSTAPXAHGVT 359
 QY 426 PAPTTTPKREPAPTTPK--KRAPTTTPKREPAPTTPK--EAPTTTPKREPAPTTPKKEPAPT 477
 Db 360 SAPDXRPXGSTAPXAHGVTSAAPDXRPXGSTAPXAHGVTSAAPDXRPXGSTAPXAHGVT 419
 QY 478 TPKETAP-----TTPKRLTTPTEKLAPTTPKEKAPPTPELIAPTTPEEPPTTPKEPAPT 533
 Db 420 SAPDXRPXGSTAPXAHGVTSAAPDXRPXGSTAPXAHGVTSAAPDXRPXGSTAPXAHGVT 471
 QY 534 TPKA-----AAPNTPKREPAPTTPKREPAPTTPKREPAPTTPKGP--APTTLKEPAP 588

DB 958 VTSAPDXRXPSTAPXAHGVTSA-----PDXRXPSTAPXAHGVTSA PD 1003

RESULT 13

US-08-479-537A-2

Sequence 2, Application US/08495537A

Patent No. 5861381

GENERAL INFORMATION:

APPLICANT: CHAMON, Pierre

APPLICANT: KIENY, Marie-Paule

APPLICANT: LATHE, Richard

APPLICANT: HAREUYEN, Mara

TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE

TITLE OF INVENTION: TREATMENT OR PREVENTION OF A MALIGNANT TUMOR

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.

STREET: P.O. Box 1404

CITY: Alexandria

STATE: Virginia

COUNTRY: United States

ZIP: 22313-1404

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/479,537A

FILING DATE: 07-JUN-1995

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: FR 90/13101

FILING DATE: 23-OCT-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: WO PCT/FR91/00835

FILING DATE: 23-OCT-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/039,320

FILING DATE: 04-APR-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/403,576

FILING DATE: 14-MAR-1995

ATTORNEY/AGENT INFORMATION:

NAME: Teskin, Robin L.

REGISTRATION NUMBER: 35,030

REFERENCE/DOCKET NUMBER: 01753-025

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 836-6620

TELEFAX: (703) 836-2021

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 2035 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

FEATURE:

NAME/KEY: Peptide

LOCATION: 128..1899

OTHER INFORMATION: /note= "The amino acids spanning

OTHER INFORMATION: /note= 1899 constitute a repeated region wherein the repeat

OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such

OTHER INFORMATION: repeats varies from 1 to 40."

FEATURE:

NAME/KEY: Peptide

LOCATION: 134

OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa

OTHER INFORMATION: Xaa Xaa which is the codon for Pro or Ala wherein Pro = CCT,

OTHER INFORMATION: CCC, CCA, or CCG; and Ala = GCT, GCC, GCA, or GCG."

FEATURE:

NAME/KEY: Peptide

LOCATION: 144

OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa

OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC,

OTHER INFORMATION: or ACG; and Asn = AAT or AAC."

FEATURE:

NAME/KEY: Peptide

LOCATION: 147

OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa

OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC,

OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."

FEATURE:

NAME/KEY: Peptide

LOCATION: 1..21

OTHER INFORMATION: /note= "Amino acids 1 to 21 are a

OTHER INFORMATION: 21 amino acid precursor sequence."

US-08-479-537A-2

Query Match 6.2%; Score 424.5; DB 2; Length 2035;

Best Local Similarity 23.7%; Pred. No. 5,3e-20;

Matches 255; Conservative 69; Mismatches 511; Indels 241; Gaps 53;

DB 140 TSLTVNKEETVETKETTITTKQSTQDCKETTSKAKETQSIKTSKAD-LAPTSTVLAKPT 198

DB 16 TVLTV-----VTSGHASSTPGEKETTSATQSSVPSSTKNAVMTSSVLSHS 65

DB 199 P-KAETTTGCP--ALTTPKEP-----TPTTPKEPASTTPKEPPTTIKSA 241

DB 66 PGSSSTTQODYTLAPATPEPASASATWQODVTSVVTBPALGSTTPPAHDVT---SAP 122

DB 242 TTPKEPAPTT-----KSAPTTPKEP-----APTTPKEPAPTTPKEPAPT 282

DB 123 --DNKPAPGSTAPXAHGVTAPDXRXPSTAPXAHGVTAPDXRXPSTAPXAHGVT 180

DB 283 TKE--PAPTT-----KSAPTTTPKEPAPTTPK-----KAPTTTPKEPAPTTPKEPAPT 329

DB 181 APDXRXPSTAPXAHGVTAPDXRXPSTAPXAHGVTAPDXRXPSTAPXAHGVT 240

DB 330 PKE--PAPTTKEP-----APTTPKEPAPTAPK-----KAPTTTPKEPAPTTPKEPAPT 376

DB 241 APDXRXPSTAPXAHGVTAPDXRXPSTAPXAHGVTAPDXRXPSTAPXAHGVT 300

DB 377 TKEPSTTPKEPAPT--TKSAPTTKEPAPT-----TKSAPTTKEPAPTTPK-----E 425

DB 301 APDXRXPSTAPXAHGVTAPDXRXPSTAPXAHGVTAPDXRXPSTAPXAHGVT 359

DB 426 PAPTTTPKEPAPTTPK-----KAPTTTPKEPAPTTPK-----EPAPTTPKAPTAKEPAPT 477

DB 360 SAPDXRXPSTAPXAHGVTAPDXRXPSTAPXAHGVTAPDXRXPSTAPXAHGVT 419

DB 478 TPKEPTAP---TTPKLTPTTPPEKLAPTTPKEPAPTTPPELAPTTPEEPPTTPEEPAPT 533

DB 420 SAPDXRXPSTAPXAHGVT---APDXRXPSTAPXAHGVTAPDXRXP---PGST 471

DB 534 TPKA---AAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKGT--APTTLKEPAP 588

DB 472 APXAHGVTAPDXRXPSTAPXAHGVTAPDXRXPSTAPXAHGVTAPDXRXPSTAPXAHGVT 530

DB 589 TTP-----KKPAKELAPTTTKEPTSTSDKPAP--TTPKGTAPTTTPKEPAPTTP 636

DB 531 TAPXAHGVTAPDXRXPSTAP--XAHGVTAPDXRXPSTAPXAHGVT---APDXR 585

DB 637 KEPAPTTPKG---TAPTTTPKEPAPTTP-----KKPAKELAPTTTKEPTSTSD 682

DB 586 PXPSTAPXAHGVTAPDXRXPSTAPXAHGVTAPDXRXPSTAPXAHGVTAPDXRXPSTAPXAHGVT 644

DB 683 KPAP--TTPK---ETAPTTTPKEPAPTTP-----KKPAPTTPPTTPPTTSEVSTP 726

DB 645 RXPSTAPXAHGVTAPDXRXPSTAPXAHGVTAPDXRXPSTAPXAHGVTAPDXRXPSTAPXAHGVT 702

DB 727 TTPKEPTT-----IKSPDESTPELSAEPPTKALENSPKEPPTTKTPAA----- 772

DB 703 DXRXPSTAPXAHGVTAPDXRXPSTAPXAHGVTAPDXRXPSTAPXAHGVTAPDXRXPSTAPXAHGVTAP 762

STREET: P.O. Box 1404
 CITY: Alexandria
 STATE: Virginia
 COUNTRY: United States
 ZIP: 22313-1404

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/083,116
 FILING DATE:

CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/479,537
 FILING DATE:

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: WO PCT/FR91/00835
 FILING DATE: 23-OCT-1991

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/039,320
 FILING DATE: 04-APR-1993

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/403,576
 FILING DATE: 14-MAR-1995

ATTORNEY/AGENT INFORMATION:
 NAME: Teskin, Robin L.
 REGISTRATION NUMBER: 35,030
 REFERENCE/DOCKET NUMBER: 017753-025
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 836-6620
 TELEFAX: (703) 836-2021

INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1867 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 FEATURE:

NAME/KEY: Peptide
 LOCATION: 128..1727
 OTHER INFORMATION: /note= "The amino acids spanning 128 to 1727 constitute a repeated region wherein the repeat is 20 amino acids, 17 of which are fixed. The number of such repeats varies from 1 to 40."

OTHER INFORMATION:
 NAME/KEY: Peptide
 LOCATION: 134
 OTHER INFORMATION: /note= "Amino acid 134 is X1 - Xaa
 OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro - CCT, CCC, CCG
 OTHER INFORMATION: or CCG; and Ala - GCT, GCC, GCA, or GCG."

FEATURE:
 NAME/KEY: Peptide
 LOCATION: 144
 OTHER INFORMATION: /note= "Amino acid 144 is Y - Xaa
 OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr - ACT, ACC, ACA
 OTHER INFORMATION: or ACG; and Asn - AAT or AAC."

FEATURE:
 NAME/KEY: Peptide
 LOCATION: 147
 OTHER INFORMATION: /note= "Amino acid 147 is X2 - Xaa
 OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro - CCT, CCC, CCA
 OTHER INFORMATION: or CCG; and Ala - GCT, GCC, GCA, or GCG."

FEATURE:
 NAME/KEY: Peptide
 LOCATION: 1..21
 OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
 OTHER INFORMATION: 21 amino acid precursor sequence."

US-09-083-116-5

Query Match 6.2%; Score 424.5; DB 4; Length 1867;
 Best Local Similarity 23.7%; Pred No. 4.8e-20;
 Matches 255; Conservative 69; Mismatches 511; Indels 241; Gaps 53;

QY 140 TSLTVNKKETTTETKETTNNKOTSDGKEKTTSAKETOSIEKTSKD-LAPTSLVLAKPT 198
 Db 16 TFLVLT-----VTSGHASSTPGGEKETSATQGRSSVSTPKMNVSMSTSLSSHS 65

QY 199 P-KAETTNGP--ALTTRKEP-----TPTRKEPASTTRKEPTPTIKSAP 241
 Db 66 PSSGSTTGGODVTLAPATEPAGSAAATGWGDVTSVPTRPALSTTPRAHDV---SAP 122

QY 242 TTPKEPAPPTT-----KSAPTTTRKEP-----APTTRKEPAPTRKEPAPT 262
 Db 123 --DNKPAPAGSTAPXAHGVTSAPOKRPXGSTAPXAHGVTSAPOKRPXGSTAPXAHGVT 180

QY 283 TRE--PAPTTT-----KSAPTTTRKEPAPTRPK---KRAPTTTRKEPAPTTTRKEPTPT 329
 Db 181 APDXRXPXGSTAPXAHGVTSAPOKRPXGSTAPXAHGVTSAPOKRPXGSTAPXAHGVT 240

QY 330 PKE--PAPTTTRKEP-----APTTRKEPAPTRPK---KRAPTTTRKEPAPTTTRKEPAPT 376
 Db 241 APDXRXPXGSTAPXAHGVTSAPOKRPXGSTAPXAHGVTSAPOKRPXGSTAPXAHGVT 300

QY 377 TKESPPTTRKEPAPTT--TSAPTTTRKEPAPT-----TSAPTTTRKEPAPTPTTK---E 425
 Db 301 APDXRP-XPGSTAPXAHGVTSAPOKRPXGSTAPXAHGVTSAPOKRPXGSTAPXAHGVT 359

QY 426 PAPTTTRKEPAPTTTRKEPAPTTTRKEPAPTTTRKEPAPTTTRKEPAPTTTRKEPAPT 477
 Db 360 SAPDXRXPXGSTAPXAHGVTSAPOKRPXGSTAPXAHGVTSAPOKRPXGSTAPXAHGVT 419

QY 478 TPKEPTAP-----TPPKTLPTTRKEPAPTTTRKEPAPTTTRKEPAPTTTRKEPAPT 533
 Db 420 SAPDXRXPXGSTAPXAHGVTSAPOKRPXGSTAPXAHGVTSAPOKRPXGSTAPXAHGVT 471

QY 534 TPKA-----AAPNTTRKEPAPTTTRKEPAPTTTRKEPAPTTTRKEPAPTTTRKEPAPT 588
 Db 472 APXAHGVTSAPOKRPXGSTAPXAHGVTSAPOKRPXGSTAPXAHGVTSAPOKRPXGSTAPXAHGVT 530

QY 589 TTP-----KKAPKELAPTTTRKEPAPTTTRKEPAPTTTRKEPAPTTTRKEPAPT 636
 Db 531 TAPXAHGVTSAPOKRPXGSTAPXAHGVTSAPOKRPXGSTAPXAHGVTSAPOKRPXGSTAPXAHGVT 585

QY 637 KEPAPTTRPKG-----TAPTLTRKEPAPTT-----KKAPKELAPTTTRKEPAPT 682
 Db 586 PXPGSTAPXAHGVTSAPOKRPXGSTAPXAHGVTSAPOKRPXGSTAPXAHGVTSAPOKRPXGSTAPXAHGVT 644

QY 683 KPAP--TPRK-----ETAPTTTRKEPAPTT-----KKAPPTTRKEPAPTTTRKEPAPT 726
 Db 645 RPXPGSTAPXAHGVTSAPOKRPXGSTAPXAHGVTSAPOKRPXGSTAPXAHGVTSAPOKRPXGSTAPXAHGVT 702

QY 727 TTRKEPTT-----THKRPDESTRELKSAEPTPKALENSPKRGGVPTTKTPAA----- 772
 Db 703 DXRPXGSTAPXAHGVTSAPOKRPXGSTAPXAHGVTSAPOKRPXGSTAPXAHGVTSAPOKRPXGSTAPXAHGVT 762

QY 773 -TKPEMTTAAKDK-----TTRDLRTTPETTAARPKMTKEATTEKTTESKITATTOVTS 828
 Db 763 DXRPXGSTAPXAHGVTSAPOKRPXGSTAPXAHGVTSAPOKRPXGSTAPXAHGVTSAPOKRPXGSTAPXAHGVT 820

QY 829 TTTQDTTPFKITLTTTLTAPRVTTTKTTTTEIMNPEETAKPKDRATNSKATTPKPO 888
 Db 821 APDXRXP-----STAPXAHGVT-----SAPDXRXPXGSTAPXAHGVTSAPOKRPXGSTAPXAHGVT 868

QY 889 KPTKAPKKTSTKPKTPRVKPKPTTPPKMTSTMTLNPSTSLAAMLOTTTRPNOT 948
 Db 869 -GSTAPXAHGVTSAPOKRPXGSTAPXAHGVTSAPOKRPXGSTAPXAHGVTSAPOKRPXGSTAPXAHGVT 911

QY 949 -PNSKLVEYNPKSEDAAGETPHMLLRPHVMEVTPDMVLYPRVNOGIIINPMLSD 1007
 Db 912 APXAHGVTSAPOKRPXGSTAPXAHGVTSAPOKRPXGSTAPXAHGVTSAPOKRPXGSTAPXAHGVT 957

QY 1008 ETNINCKRPVGLTTL-LANGTLVAFRGHYFMWLSFSPSPARITTEYVWGIIPSPID 1062

APPLICATION NUMBER: WO PCT/FR91/00835
 FILING DATE: 23-OCT-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/039, 320
 FILING DATE: 04-APR-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/403, 576
 FILING DATE: 14-MAR-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Teskin, Robin L.
 REGISTRATION NUMBER: 35, 030
 REFERENCE/DOCKET NUMBER: 017753-025
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 836-6620
 TELEFAX: (703) 836-2021
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1867 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 FEATURE:
 NAME/KEY: Peptide
 LOCATION: 128..1727 /note="The amino acids spanning 128 to 1727 constitute a repeated region wherein the repeat is 20 amino acids, 17 of which are fixed. The number of such repeats varies from 1 to 40."
 OTHER INFORMATION:
 OTHER INFORMATION:
 FEATURE:
 NAME/KEY: Peptide
 LOCATION: 134 /note="Amino acid 134 is X1 = Xaa which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCG, CCA or CCG; and Ala = GCT, GCC, GCA, or GCG."
 OTHER INFORMATION:
 OTHER INFORMATION:
 FEATURE:
 NAME/KEY: Peptide
 LOCATION: 144 /note="Amino acid 144 is Y = Xaa which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACG, AAT or AAC."
 OTHER INFORMATION:
 OTHER INFORMATION:
 FEATURE:
 NAME/KEY: Peptide
 LOCATION: 147 /note="Amino acid 147 is X2 = Xaa which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCG, CCA or CCG; and Ala = GCT, GCC, GCA, or GCG."
 OTHER INFORMATION:
 OTHER INFORMATION:
 FEATURE:
 NAME/KEY: Peptide
 LOCATION: 1..21 /note="Amino acids 1 to 21 are a 21 amino acid precursor sequence."
 OTHER INFORMATION:
 OTHER INFORMATION:

Query Match	6.2%;	Score 424.5;	DB 2;	Length 1867;
Best Local Similarity	23.7%;	Pred. No. 4.8e-20;		
Matches 255;	Conservative 69;	Mismatches 511;	Indels 241;	Gaps 53

OY	140	TSLTVNKETVELKETTNNKOTSDGKEKITTSAKETOSIETSIAKO-LAPTSKYLAPKT	198
Dd	16	TLVAV-----VTSGHASSTPGGEKETSAORSSVPSESTENNAVSMTSSVLSSMS	65
OY	199	P-KAETTTKG-P-ALTPEKP-----TPTPKPEASTTKPEPTPTTIKSAP	241
Dd	66	PGSGSITTOGDVLTALPATERASSAATWOGDVTSVPTVRPALGSTPPAHADVT---SAP	122
OY	242	TTTTKEPAPTTT-----KSAPTTPEKP-----APTTEKEPATTPKEBAPT	282
Dd	123	--DNKPAGGSTAPXAHGVTSAPDXRPBSTAPXAHGVTSAPDXRXPBSTAPXAHGVTS	180
OY	283	TRE--PAPPTT-----KSAPTTPEKAPPTPK-----KPAPTPEKAPPTPKPTPT	329

```

Db 181 APDXXRPGSTAPXAHGVTASDDXRPXGOSTAPXAHGVTASAPDXXRPGSTAPXAHGVTAS 240
Qy 330 PKE--PAPTTKEP-----APTTKEPAPTA PK---KRAPTTKEPAPTTTKEPAPTT 376
Db 241 APDXXRPGSTAPXAHGVTASDDXRPXGOSTAPXAHGVTASAPDXXRPGSTAPXAHGVTAS 300
Qy 377 TKEPSPTTKEPAPTT--TKSAPTTTKEPAPTT-----TKSAPTTPEPSPTTTK-----E 425
Db 301 APDXXR--XPGSTAPXAHGVTASDDXRPXGOSTAPXAHGVTASAPDXXRPGSTAPXAHGVT 359
Qy 426 PAPTTTKEPAPTTTTPK---KRAPTTKEPAPTTPK---EPAPTTTTPKRAPTAKEPAPT 477
Db 360 SAPDXXRPGSTAPXAHGVTASDDXRPXGOSTAPXAHGVTASAPDXXRPGSTAPXAHGVT 419
Qy 478 TPKETAP-----TTTKKLITPTTPEKLAPTTPEKAPPTTPEELAPTTPEEPTTTTPEEAPT 533
Db 420 SAPDXXRPGSTAPXAHGVTAS---APDXXRPGSTAPXAHGVTASAPDXXRPGSTAPXAHGVT 471
Qy 534 TPXA---AASNTPEKAPPTTPKBPAPTTTPEKAPPTTPEKAPTTTPEKAPTTTPEKAP 588
Db 472 APXAHGVTASAPDXXRPGSTAPXAHGVTASAPDXXR--XPGSTAPXAHGVTASAPDXXRPGS 530
Qy 589 TTP-----KKRAPKELAPTTTKEPSTSDXAP--TTPGATAPTAPEKAPPTT 636
Db 531 TAPXAHGVTASAPDXXRPGSTAP--XAHGVTASAPDXXRPGSTAPXAHGVTAS---APDXX 585
Qy 637 KEAPPTTPKG---TAPTTTKEPAPTTT-----KKRAPKELAPTTTKGPTSTSD 682
Db 586 PXPGSTAPXAHGVTASAPDXXRPGSTAPXAHGVTASAPDXXRPGSTAP--XAHGVTASAPDX 644
Qy 683 KRAP--TTPK---ETAPTTTKEPAPTT-----KKRAPTTPEPTPTTSEVSTP 726
Db 645 RXPXGSTAPXAHGVTASAPDXXRPGSTAPXAHGVTASAPDXXRPGS--TAPXAHGVTAP 702
Qy 727 TTTTKEPTT-----IHXSPDESTPELSABPTPKALENSPKEPVTTPPA----- 772
Db 703 DXRXPXGSTAPXAHGVTASAPDXXRPGSTAPXAHGVTASAPDXXRPGSTAPXAHGVTAP 762
Qy 773 --TPEMTTAKOK--TTERDLITPTETTTAPAKMTKETATTTTEKTESKITATTTOVTS 828
Db 763 DXRXPXGSTAPXAHGVTASAPDXXRPGST--APXAHGVTASAPDXXRPGSTAPXAHGVTAS 820
Qy 829 TTTQDTTPEKITTLKTTTLAPAVTTTTKTITTTTILMKPEETAKPKDRATNSKATTPKQ 888
Db 821 APDXXRPGS-----STAPXAHGVT-----SAPDXXRPGSTAPXAHGVTASAPDXXRXP- 868
Qy 889 KPTKAPKKTSTTKKPKTTPRVKPKTTTPPKMTSTWPELNPSTRIAEMLQTTTPNOT 948
Db 869 -GSTAPXAHGVTASAPDXXR--XPGSTAPXAHGVTAS--PDXRPX-----PGST 911
Qy 949 -PNSKILVEVMPKSEDDGAGEGTPHMLLRPHVFMEPTVPMQDYLPRVNOGIIINPMLSD 1007
Db 912 APXAHGVTASAPDXXRPGSTAPXAH-----GVTASAPDXXRXPXGSTA---PXAHG 957
Qy 1008 ETNINCKRPVDGLTT--LRNGTLVAERGHFWMLSPFSPSPARKITTEWQIIPSPID 1062
Db 958 VTSAPDXXRXPXGSTAPXAHGVTAS-----PDXRPXGSTAPXAHGVTASPD 1003

RESULT 12
US-09-083-116-5
; Sequence 5, Application US/09083116
; Patent No. 6203795
; GENERAL INFORMATION:
; APPLICANT: CHAMBOB, Pierre
; APPLICANT: KIENY, Marie-Paule
; APPLICANT: LARHE, Richard
; APPLICANT: HAREUVENT, Mera
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
; TITILE OF INVENTION: TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.

```



```

OY      696 TKPEP-----APTPKKRAPPTPPPTTSSEVSTTTTKEPTTHKSDSESTPELSAE 749
          | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db      650 LPPIQKGNNIPNMLPENPSDSEVEYPRPDNDGNSNNMKSKNNT---PNEFIPSPGDN 706
          | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
OY      750 PTPKALENSPK-----EPGVPTTK 768
          | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db      707 PYKGHERIKRPHRSNDIYYDNNVNKNKNDDEPIPPANNE 745
          | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |

RESULT      8
PCT-US92-00018-2
Sequence 2, Application PC/TUS9200018
GENERAL INFORMATION:
APPLICANT: Hoffman, Stephen L.
APPLICANT: Charoenvitl, Yupin
APPLICANT: Hedstrom, Richard
APPLICANT: Khumstich, Sirisin
APPLICANT: Rogers IV, William O.
TITLE OF INVENTION: Protective malaria sporozoite surface protein
TITLE OF INVENTION: immunogen and gene encoding
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: A. David Spevack
STREET: NMDC Building 1 T-12 National Naval
STREET: Medical Center
CITY: Bethesda
STATE: MD
COUNTRY: USA
ZIP: 20814-5044
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/00018
CLASSIFICATION: 424
FILING DATE: 19920103
ATTORNEY/AGENT INFORMATION:
NAME: Spevack, Avram D.
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 295-4033
TELEFAX: (301) 295-6759
INFORMATION FOR SEQ ID NO.: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 826 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US92-00018-2
```

```

Db      493 PS-----NPNESNNEESNPNNE--PSNPKK-----PSNPNNE--P 523
QY      517 TTPEEPTPTTPEEPAPTTPKAAAPNPKPE-APTTEKEPAPTTPKEPAPTTPKEPTAP 575
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      524 SNPNNE--PLNPNNEB-----SNPNESNNEESNPNNEE--PSNPKX--PSNPNNE----- 564
QY      576 KGTAPTTLKEPAPTTPKPKPAKELAPTTTKEPTSTSDKPAPTTPKGTAPTTPKEPAPT 635
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      565 -----PSNPEEPNPEE--PSNPKBP-----SNPEEPIINPELINKEESNNEESN 606
QY      636 PKEPAPTTPKGTAPTTLKEPAPTTPKPKPAKELAPTTTGPSTSTSDKPAPTTPKETAPT 695
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      607 PKEPI-----NPNESNPKBP-----INPEEDNEPDLIODEPIEPRNDSNVIAI 649
QY      666 TPKEB-----APPTPKKPAPTTPEPTPTTSEVSTPTTKEPTTIHKSDESTPELSAE 749
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      650 LPPIQKGNNTPSNLPEPNPSDEVEYERPRNDNGNSNNTKSSKNKT--PNEPIPSGDN 706
QY      750 PTPKALENSPK-----EPGVPTT 768
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      707 PTKGHEEIRPKPHRSNDYIVUDNNVANKNNDDEPIIPNNE 745

```

```

RESULT          9
5202236-13
; Patent No. 5202236
; APPLICANT: MATCH, KATHY J.; ANDERSON, DAVID M.; STRAUSBERG,
; SUSAN L.; MCCANDLISH, RUSS-WEL, TENN; FLIPULA, DAVID
; TITLE OF INVENTION: METHOD OF PRODUCING BROADHESIVE
; PROTEIN
; NUMBER OF SEQUENCES: 39
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/528,762
; FILING DATE: 25-MAY-1990
; APPLICATION NUMBER: 82,456
; FILING DATE: 07-AUG-1987
; APPLICATION NUMBER: 933,945
; FILING DATE: 24-NOV-1986
; APPLICATION NUMBER: 650,128
; FILING DATE: 13-SEP-1984
; SEQ ID NO:13:
; LENGTH: 652
5202236-13

```

Query Match	7.08;	Score 476.5;	DB 6;	Length 652;
Best local Similarity	29.78;	Pred. No. 5.1e-24;		
Matches 221;	Conservative	284;	Indels 169;	Gaps 45

Query Match	7.2%	Score 488.5	DB 5:	Length 826:
Best Local Similarity	26.4%	Pred. No. 11e-24:		
Matches 153;	Conservative	54;	Mismatches 223;	Indels 149;
				Gaps 34;
Qy	233	PTPTTKSAPPTTKEPA-PTTTSAPPTTKEPA-PTTKEP---	APTTPKEP-APTTPKE	285
	11	:::::	:::::	
Db	273	TTCKVNDQDPIPLPYIPKIKPEKSPNEBPVNDPNDPNNNNNNNNNNPNNPN		332
Qy	286	PADPTTKSAPPTTKEP-APTTPKEP-APTTPKEP-APTTPKEP-PTTTEKEPA	PTTTEKEPA	341
	11	:::::	:::::	
Db	333	PNNNNNNNNNNNNPNNPNNPNNPNNPNNPNNNNNNNNNDSSNPNHKK		392
Qy	342	PTTTEKEPAPTAPK-KEAPTTPKEPAPTTTEKEPAPTTTEKEPSTTTEKEPA	PTTTSAPT	399
	11	:::::	:::::	
Db	393	KRNPKRNPKPKPKPKPKPNPNESNNKNNPN----	ESNNKPNP----	NEPSN 443
Qy	400	TKEDAPTTPKSAPTTKEPSPTTTEKEPAPTTTPKEP-PTTPKAP-APTTPKEA	APTTPKEA	457
	11	:::::	:::::	
Db	444	PNKNPNP-----NESPNNKPNPN-----EPINPNEPSNPNDSNPNASPNNE-	PSNPNP	492
Qy	458	PADPTTKSAPPTTKEP-APTTPKEPAPTTPKKTPTTTEKEPAPTTTEKEA	APTTPKEA	516

[illegible]

```

QY 415 PKE-----PSPPT-----422
DB 407 TSETESVIRDEMCWLEKNGECAKATVGVIGDKRIENGMATMIPNDTHFRPK 466
QY 423 -----TKEP-----APTPKE 433
DB 467 VADVGNITISRCRKAGKLEPRDRSLDFTIPVAGHNSCSIIIVGSGDKIHVSPGSD 526
QY 434 PAPTTPKRPAPTTPKE-----PAPTTPKRPAPTTPKRPAPTTPK 475
DB 527 VSLIS---APIQSELEFNEVYCDCTAKYGAHSGYOTSADPVTYTTAKPPTTT--TGA 580
QY 476 PTPKETATPTTKKLPTTPPEKLAFTTPPEKRAPTPPEELAPTPPEPTTPPEEPAPTTP 535
DB 581 PQQPTTYYTGSFSPKPTTYYTKATTTT-----TLNPLITTTTQOKPPTTT-----TT- 627
QY 536 KAAANPTPREPAPTTPKEPAPTTPKEPAPTTPKATPTTKATPTTKERAPTTPK 595
DB 628 --KVGCKPPIATTTTTLKPIYTTTTKATTTTTPVPTT---TTTKRDEMPTTTTTP 680
QY 536 PK---ELAPTTTKBPTS-----TTSKRAPTPPKG--TAPTTPKEPAPTTP 635
DB 681 PDIGDITPTPIEKMLQKTYRMIVDYSGLLSDNDSEPIGSOAGQIADTSLNLEPVQTH 740
QY 636 PKEPAPTTPKGTAPTTTKE---PAPTTPKAPKELAPTTTKGPTSTTSKRAPTPPKET 692
DB 741 KSTGLPIDPMVGLPDPKSGMLVHPYTNQTMGSLSVSYLAKNLTVDETYG--LPIDT 798
QY 693 APPTPKPAPTTPKRPAPTTPETPTPTTSEVSTPT-----TTKETPTTHKS----- 738
DB 799 LTGVPIDVSLIPFN--PETSELEDPSIDELMNGTIAGIYSGISASELSOKALIDPA 856
QY 739 -----PDEST-----PELSAEPTPKALENSPK 760
DB 857 TNNVYGEFGLINPATGVMIRPFLGSEHQFSPSEIEDGGIIPREVAANDKRLSLP- 915
QY 761 EPGVPTTKTPAATKPEMTTAKDKITER-----DURTPPETTTAAPKM 803
DB 916 -PSPF-----ESIPF-----KDKKIDISIELMADIESGRLIGVSKRPIFGSIAGDINP 963
QY 804 TKETATTEKTESKITATP--TOVTSSTTODTTPFKITTLTKTTTLAKVTTTKKITTTE 862
DB 964 IMKPTOTDSVYKRPIDPTGFLPFPNPTGLHINLTNNNTDSSAAGVAKVAVSGITDN 1023
QY 863 IMNKP--EETAKPKD-----RATNSKATTPKQAPTKAPKPTSTKKKKTMP----- 907
DB 1024 VYGLPVEITGLPKDPSGDIPEFNSTGELVDPSTGKPIINNSTAGIYSGKGLPIEDENG 1083
QY 908 -----RVRKFKTTPTPRKMTS--TMPE-----927
DB 1084 NLFDPSTNLPIDGNOLVNPETNSTVSGTSTTKPKPGIIPVNGGCVVPDEEAKDQADKG 1143
QY 928 -----LNPTRIAEAMLOTTTRPNQTPNSKLVENPKSEDAAGEETPHMLRPHVFP 982
DB 1144 KQGLIVPTINSINKDPNTNQYSNTGNI---INP--ETGKVIYPSGLPSLAVPSFNP 1197
QY 983 EYTPD-----MDYLPVPNOGIIINP-----MLDENYICNGK-----1015
DB 1198 QQDDELIGKVDVYVYGLPIYPSSTGEIIDPATKLPISGAVAGDEILVINTTTDEVYGLP 1257
QY 1016 -----PYDGLTTLRNGTLV 1029
DB 1258 IDELETGLPRDPVSGLPOLPNGTLV 1281

```

```

RESULT 7
US-07-638-431-2
Sequence 2, Application US/07638431
Patent No. 5198533
GENERAL INFORMATION:
APPLICANT: Hoffman, Stephen L.
APPLICANT: Charoenvit, Yupin

```

```

APPLICANT: Hedstrom, Richard
APPLICANT: Khushmith, Srisin
APPLICANT: Rogers IV, William O.
TITLE OF INVENTION: Protective malaria sporozoite surface protein
TITLE OF INVENTION: Immunogen and gene
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESS: A. David Spevack
STREET: NMDC Building 1 T-12 National Naval
STREET: Medical Center
CITY: Bethesda
STATE: MD
COUNTRY: USA
ZIP: 20814-5044
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/638,431
FILING DATE: 19910110
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Spevack, Avrom D.
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 295-6759
TELEFAX: (301) 295-4033
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 826 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-638-431-2

```

Query Match 7.2%; Score 488.5; DB 1; Length 826;

Best Local Similarity 26.4%; Pred. No. 1,1e-24;

Matches 153; Conservative 54; Mismatches 223; Indels 149; Gaps 34;

```

QY 233 TPTTISAPTTKEPA-PPTTKSAPTTKEPA-PPTTKP---APTTPKEP-APTTPKE 285
DB 273 TTKVNDQDIPPIVPIPKIPEKPSNBEVPVNDPNPNPNPNPNPNPNPNPNPN 332
QY 286 PAPTTPKSAPTTPKEP-APTTPKEP-APTTPKEP-APTTPKEP-PTTTPKEPAPTTPKE 341
DB 333 PNNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNH 392
QY 342 PTPPKPAPTTPK--KAPTTPKEPAPTTPKEPAPTTPKEPSTTPPKPAPTTPK 399
DB 393 RRPKRRNNPNKPKPNKPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPN 443
QY 400 TKEPAPTTPKSAPTTPKEPSTTPPKPAPTTPKEPA-PPTPKP-APTTPKEPAPTTPKE 457
DB 444 PKNPNP---NEPSNKNKNPN---EPLNPNESNPNPNPNPNPNPNPNPNPNPN 492
QY 458 PAPTTPKRPAPTTPKAP-APTTPKETAPTTPKLLPTTPPEKLAFTTPPEKAPTTP 516
DB 493 PS-----NPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPN 523
QY 517 TTPPEPTPTTPPEPAPTTPKAAAPTTPKEP-APTTPKEPAPTTPKEPAPTTP 575
DB 524 SNPNP--PLNPNP-----SNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPN 564
QY 576 KGTAPTTLKEPAPTTPKAPKELAPTTTKETSTTSKRAPTTTKGAPTTPKAP 635
DB 565 -----PSNPEPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPN 606
QY 636 PKEPAPTTPKGTAPTTLKEPAPTTPKAPKELAPTTTKGPTSTTSKRAPTTTPKETAPT 695
DB 607 PKEPT-----NPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPN 649

```


QY 477 -----TTPKETA-PTTP-----KKL 490
Db 842 AGQIADTSMLEFVQTHKSTGLPIDMWGLPDPKSGNLVHPTNQTMSGLSYLAANKL 901
QY 491 T-----PTTPEKLAPTTP-----504
Db 902 TWDQDETGLPIDLLGLYLDPLPILPFENPETGELFDPISDEIMNGTLAGIVSGISASES 961
QY 505 ----KRAPTPPE-----LAPTT-----PEEPPTTP--DEAPPTPKAA 539
Db 962 LLSQKAPRIDPATNMVGEGLINPATGVMTPGSLGPSEQTPFSEIIDGGIIPPEVAA 1021
QY 540 PNTKPEAPPTTKEPAATPK-----PA 563
Db 1022 ANADKRLSPESVPSIEKDKIDISGLMYDIESGRILIGVSKRPIPSIAGDLNPI 1081
QY 564 PTPPKETAPPTKGAAPTLKEPAATPKPKAPKELAPTTKEPTSTSDKRAPTPPKGT 623
Db 1082 MKTPGTQDSVTKRIPPTT---GLPFNP--PTGHLINPNNNTMDSSEAGAYVAVSNGI 1136
QY 624 APPT-----KPEAPTPKPEAPPTPKGTAPTLKEPAATTP-----KKPAK-- 666
Db 1137 KTDNVYGLPVDLITGLPKDVSQDIPFNSTGELVDSGTGKPINNTAGIYSGKRLPPIE 1196
QY 667 ----ELAPTTTGP-----TSTT-----SDKAPPTPKETAPTTKEPA--- 701
Db 1197 DENGNLDPSTKLPIDGNQLVNPENSTVSGSTSGSTGKPKGPIPVNGGVVDEDEKDO 1256
QY 702 -----PTPKKAPATTPPETPTPTSEVSTPTTKE--PTTIHKS--PDESTPEL 746
Db 1257 ADKKGDLIVPINSINKDPVTNTQYNTGNININPETGKVIQSLGSLNYSFNPOQ 1316
QY 747 SAEPTRKALE---NSPKPEGVPTTKTPAATKPEMTTAKDTERDILRTTPETTAAPK 803
Db 1317 TDELITGKRVQDVGLPYDPSGELIIDPATKLPISVAGSEILILEVNTITDEVTLGP-I 1375
QY 804 TKEPATTEKTESKITATTQVSTTQDPTTPKITTILKTTLAPVTTKKTITTE- 862
Db 1376 DLEGLPDRP-----VSGLPOLNGLVND-----PSKKRIIPSGHS 1411
QY 863 --INMKPEETAKPKDRATNSKATTPKPKP--TKAP--KKPTSTKPKTMPRVRKPKT-- 914
Db 1412 GFINTSGEOSHEDPSG-----KPLDPNTGLPDEDSSGLINPETGDKLQSHSGTF 1465
QY 915 TTPPKMTSTMPELNPSRIAPAMIQTTTRNQTNPNSKLVANR--SEDA----- 963
Db 1466 MPVPGKPGENGIMTPQOILEAL-----NKLPTSEVNISPPSSDAVDPDRPTNTW 1518
QY 964 -----GGAEGETPHML--RPHVMEVTPDMQYLPRVNOGIII 1001
Db 1519 NKISQOTQVDGKTIIPGSAASVHITAGTPTQIDPTTGLP--SDPSTGLPFIFGENVLY 1576
QY 1002 NEMLSL-----ETNIC-----NGKPYDGLTTL--RNGLYV-AF 1031
Db 1577 DQGTQBOJIKGSVPYSLVYKERNIVTEAAYGLPVDKGTGFPIDPSIYLPAKNGELDPT 1636
QY 1032 RGHYF 1036
Db 1637 SGKYF 1641

RESULT 3
5202236-25
; Patent No. 5202236
; APPLICANT: MAUGH, KATHY J.; ANDERSON, DAVID M.; STRAUSBERG,
; SUSAN L.; MCCANDLISS, RUSS; WEI, TENA; FILIPULA, DAVID
; TITLE OF INVENTION: METHOD OF PRODUCING BIOADHESIVE
; PROTEIN
; NUMBER OF SEQUENCES: 39
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/528,762
; FILING DATE: 23-MAY-1990
; APPLICATION NUMBER: 82,456

FILING DATE: 07-AUG-1987
APPLICATION NUMBER: 933,945
FILING DATE: 24-NOV-1986
APPLICATION NUMBER: 650,128
FILING DATE: 13-SEP-1984
SEQ ID NO: 25
LENGTH: 744
5202236-25

Query Match 7.4%; Score 506.5; DB 6; Length 744;
Best Local Similarity 29.9%; Pred. No. 6.2e-26;
Matches 239; Conservative 78; Mismatches 308; Indels 175; Gaps 50;

QY 116 PKITAKINRPSLPRMSDTSKESLSLVNKEETVETKETTTTKQSTDSKEKTSAKE 175
Db 24 PKMTYPPYKPKPSYP-----YKSKPY-----KPKIT----- 53
QY 176 TQSIKTSADLAPTSKYLAKPKEATTTKQALTPPKPEPTTPPKPE---PASTTPKE 231
Db 54 -----YPPYK--AKPS-----YPPYKPKKTYPPYKPKLTPPYKPKP 92
QY 232 PPTPKISAPT--TKEPAPTTKSAPTTKPEAPTTKEPAATTPKPEAPTTKEPA--P 288
Db 93 SYPTPKSKPYKPKITYPPYKAKPSYPPYKPKPKTYPP--YKPLTYPPYKPKASYP 151
QY 289 TTTKSAPTTPKPEAPTTPKPAAT--TKEPAPTT-----KPEPTTPKPEAPTTKEP--- 340
Db 152 PLYKPKPSYP--PSYTKTKITYPPYKPKLTPPYKPKPSYPPYKPKPKTYPPYKPKLT 209
QY 341 --APTTKEPA--PTAKKPA--PTTPKPEA--PTTPKPEA--PTTPKES--PTTPKPEA 389
Db 210 YPPTYKAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPS 269
QY 390 -PTTSAPTTTKEPAATTTKSAPTTPKPSPTTKEPA--PTTPKPEA--PTTPKPK 442
Db 270 YPSTYKAKPSYT--YPSYKAKPYT-----PYKAKPSYPPYKAKPKTYKAKPK 322
QY 443 A--PTTPKPEA--PTTPKPEA--PTTPKPA--PTAKPEAPTTKETAAPTTPKLTPT 494
Db 323 SYPTPKAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPS 382
QY 495 PEKLAPTPE--KRAPTTPELAPTTPEPT--PTPEEPA--PTPKA--AAPNT-----P 543
Db 383 TYKAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPSY 438
QY 544 KEPAATTPKPEAPTT-----KPEAPTTKETAAPT--PKGAPTTLKEPAATTPKPAK 597
Db 439 SYPTPKPKISYPPYKAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPSYPPYKAKP 498
QY 598 ELAPTTKEPTSTSDKRAPTTKGTAPTTKPEAPTTTKEPAATTPKGTAPTTKEPA- 656
Db 499 SYPTPKAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPS 546
QY 657 -PTPKKPAPELAPTTTSGPT--STSDKPAATTPKETAAPTTPKPEAPTTPKPA--PT 711
Db 547 PPTKAKPYT--AKPTYKAKPYTSTYKAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPS 601
QY 712 ---TPEPTTPPTSEVSTPTTKE---PTTIHKSPT-----ESTPELSAAPT--PKALENS 758
Db 602 YKAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPSYPPY 661
QY 759 PKEGVPTTKTPAA-----TKPEMTTAKDKTERDLRTTPETTAA-----KMT 804
Db 662 PSYP--PTYKAKPSYPPYKAKPSYPPYKAKPKTYPPYKAKPSYPPYKAKPSYPPY 719
QY 805 KETATTEKTESKITATT 824
Db 720 YKAKPTYSTYKAKPYST 739

RESULT 4
US-09-103-429A-3

Matches 226; Conservative 33; Mismatches 211; Indels 342; Gaps 40;

QY 190 TSKVLAKPTTKAETTTGKPAITTPKEPTTPPKPAETTPKKEPTTPKSAF-----TTP 244
 DB 9 TALGLVAAAREVSDAEKNPALHEPHDPX--PAEQXKLLPEKEDCKFYCEGLKFLIAP 66
 QY 245 KKPAPTT-----TKSAPTTPKKPAITTPKPAETTPKKEPTTPKKEPTTPK 292
 DB 67 KDAAPGTEEFESAGTQVHAALACCTLPKPAETTT--QAPATQO--APTITQ--APTITQ 119
 QY 293 SAPTTKPAETTPKPAETTPKKEPTTPKKEPTTPKKEPTTPKKEPTTPKKEPTTPK 352
 DB 120 QAPTTTQATTTT-----QAPTTTQATTTT-----QAPTTTQATTTT----- 156
 QY 353 PKKPAETTPKKEPTTPKKEPTTPKKEPTTPKKEPTTPKKEPTTPKKEPTTPKSAF 412
 DB 157 ---QAPTTTQ--APTITQ--APTITQ-----APTITQ--APTITQ--APTITQ 198
 QY 413 TPKKPAETTPKKEPTTPK--KPAETTPKPAETTPKKEPTTPK----- 456
 DB 199 TTPKPAETTPK--TPAETTPKPAETTPKPAETTPKPAETTPK----- 253
 QY 457 ----- 464
 DB 254 IHLHPDKYCNLFYQCSNGYTFEQRCRPEGLFNPYVQRCDSPANVECDGEISPAVPYE 313
 QY 465 K-----PA----- 467
 DB 314 GNEDEDIDIGLLDNGCPANFEIDWLPHGNCRDKYQCVHGNLVERGAGTHFELQ 373
 QY 468 -----TPAP--KPAETTP-----TPRET 482
 DB 374 QCDHIELVGCILPGSESEVDDEDACTGWCFTPEIEMELPNCGRPADSFIDHLLPHES 433
 QY 483 -----APTTPKKLTPTT-----PEKLAETTPKKEPTTP 508
 DB 434 DCGOYLCVHGOTARPCRGMLHSPATOCSESVTACQVFECDSDMQSTAPAPAP 493
 QY 505 TTPKPAETTPKKEPTTPKKEPTTPKKAAPNTPKKPAETTPKKEPTTPKKEPTTPK 568
 DB 494 TAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAP 548
 QY 559 ETAPTT--PKGTAETTPKKEPTTPKPKKPAKELAPTTTKEPTTPKSDKPAETTPK 625
 DB 549 PESPTTTPKPAETTPK--APT-----AVPEIPIVTSAPTAAP--AAPTAAPTAAP 598
 QY 626 TTPKPAETTPKKEPTTPKKAETTPKKEPTTPKKEPTTPKKEPTTPKKEPTTPK 684
 DB 599 TTVAPETPT-----VTSPTAAPT--AAPAPNT-----TVTVPTAAPTAP 641
 QY 685 APTTPKPAETTPKKEPTTPKKAETTPKKEPTTPKKEPTTPKKEPTTPKKEPTTPK 744
 DB 642 AP-----NTTVAPPTAAPT--AAPAPNTTVTVPTAAPTAP-----PTVAH-----AP 685
 QY 745 ELASPTPKALENSPKRPGVPTTKPAATKPE 776
 DB 686 NTTAAP-----VTTTSAPAPT--PE 703

RESULT 2
 US-08-928-361B-5
 ; Sequence 5, Application US/08928361B
 ; Patent No. 6071518
 ; GENERAL INFORMATION:
 ; APPLICANT: Petersen, Carolyn
 ; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
 ; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
 ; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
 ; NUMBER OF SEQUENCES: 30
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESS: PETERS, VERNY, JONES & BIKSA
 ; STREET: 385 Sherman Avenue, Suite 6

CITY: Palo Alto
 STATE: CA
 COUNTRY: USA
 ZIP: 94306-1840
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: IBM PC compatible
 SOFTWARE: Patentln Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/928,361B
 FILING DATE: 12-SEP-1997
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 60/026,062
 FILING DATE: 13-SEP-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: VERNY, HANA
 REGISTRATION NUMBER: 30,518
 REFERENCE/DOCKET NUMBER: 480.76-1(HV)
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 650-324-1677
 TELEFAX: 650-324-1678
 INFORMATION FOR SEQ. ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1837 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-928-361B-5

Query Match 7.5%; Score 508; DB 3; Length 1837;
 Best Local Similarity 21.3%; Pred. No. 1,4e-25;
 Matches 295; Conservative 100; Mismatches 494; Indels 496; Gaps 48;

QY 98 KVTPTSTTQHNKVKSTSKRTTAKPINRSLPSPDSKETSLSLVKETTVEKETTT 157
 DB 307 KHTTT 366
 QY 158 TNKQSTDSGKEKTSKESKOSIEKTSANDLAPTSVKLAKPTPKAETTPKPAETTPK 217
 DB 367 TTT 426
 QY 218 PTPKPAETTPKKEPTTPKKEPTTPKKEPTTPK--APTITTSAPTTPKPAETTPK 275
 DB 427 TTT 486
 QY 276 KKPAPTTTPKPAETTPKSAPTTPKKEPTTPKKEPTTPKKEPTTPKKEPTTPK 324
 DB 487 TTTATTTTTTKKPTT 546
 QY 325 -----PTPTT-----PK 331
 DB 547 AGGATVGVIGKGRILENGMAFTMIPNDTHVRFKVKVDGNTISVRCGAGKLEPPD 606
 QY 332 EPAPTTKEP-----APTTPKEPA--PTAPKRP----- 356
 DB 607 RSLDFTIPVAGHNSCSIIIVGSGGKIHVSPYSGKSVLSAIDIQCELENEYCDTCT 666
 QY 357 -----APTTPKEPAPT--PKKPAETTPKKEPTTPKKEPTTPKKEPTTPK 395
 DB 667 AKYGAHSGYOTADPVTYTTAKPTTTTTTGACQPTTTTOSPSKPTTTTTTKATTTTTT 726
 QY 396 -----APTTPKEPAPTTPKSAPTTPKKEPTTPKKEPTTPKKEPTTPKKEPTTPK 451
 DB 727 LNPITTTTQKPTTTTTTKVGVKRPVIAITTTTLKPIVITTTTKATTTTTTTTVPPT 781
 QY 452 PTPKPAETTPKKEPAP-----TAPKEPAP----- 476
 DB 782 TTTTKRDEMTTTTPLDIDGIDETITPIPIEKMLDKYRMIYDYSGLLSDNDEPTIPGSO 841

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:25:33 : Search time 61.21 Seconds
(without alignments)
466,904 Million cell updates/sec

Title: AA3
Perfect score: 6814
Sequence: 1 MAWKTPYLLILSLVFIQ.....ARATTRSGTSLSKWTNCP 1270

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*

- 1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
- 2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
- 3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
- 4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
- 5: /cgn2_6/ptodata/2/1aa/PTCTUS.COMB.pep:*
- 6: /cgn2_6/ptodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

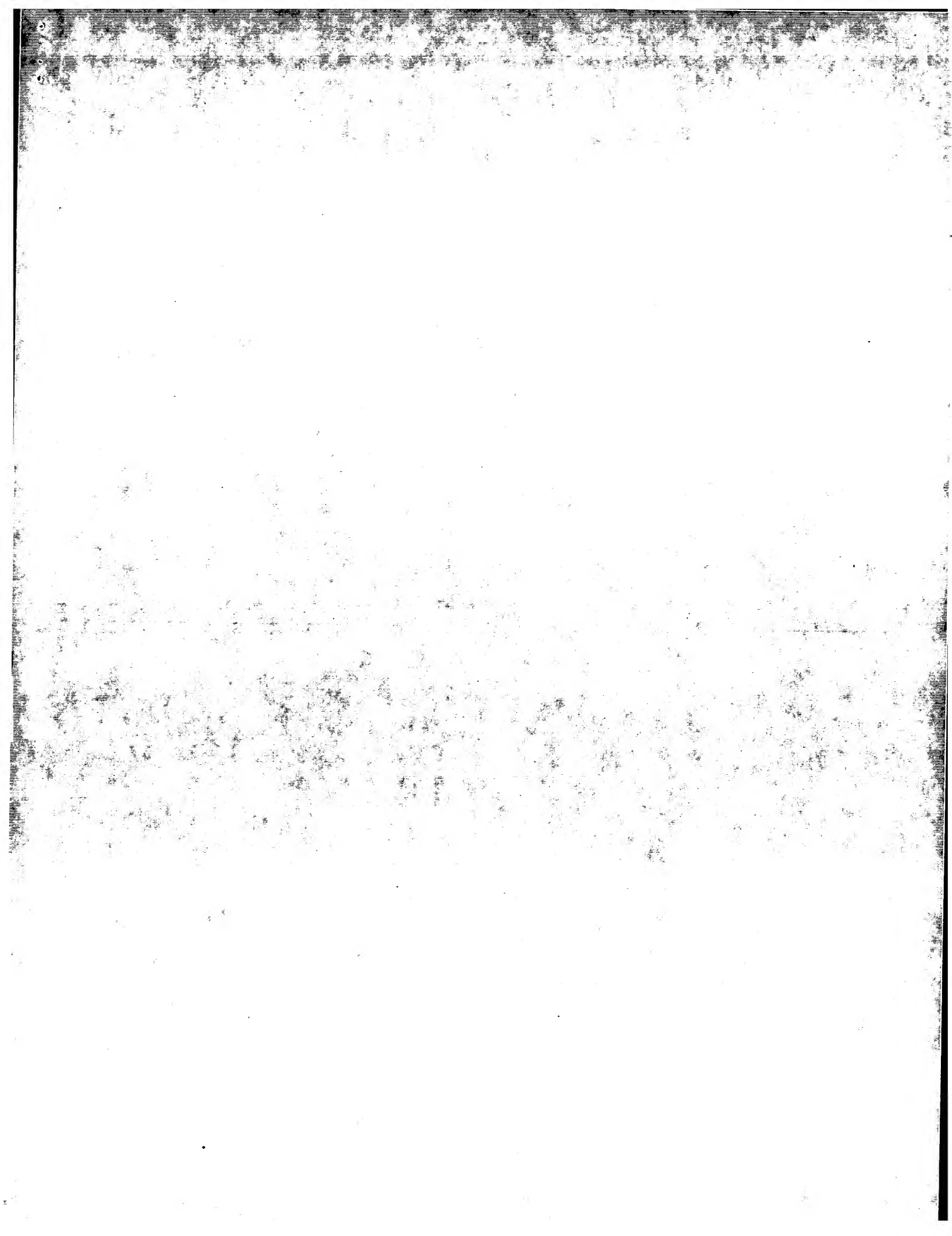
SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	522	7.7	805	US-09-103-429A-4	Sequence 4, Appli
2	508	7.5	1837	US-08-928-361B-5	Sequence 5, Appli
3	506.5	7.4	744	5202236-25	Patent No. 5202236
4	498.5	7.3	786	US-09-103-429A-3	Sequence 3, Appli
5	489	7.2	1721	US-08-700-651-5	Sequence 5, Appli
6	489	7.2	1721	US-08-928-361B-6	Sequence 6, Appli
7	488.5	7.2	826	US-07-638-431-2	Sequence 2, Appli
8	488.5	7.2	826	PCT-US92-00018-2	Sequence 2, Appli
9	476.5	7.0	652	5202236-13	Patent No. 5202236
10	452	6.6	960	US-09-219-849-5	Sequence 5, Appli
11	424.5	6.2	1867	US-08-479-537A-5	Sequence 5, Appli
12	424.5	6.2	1867	US-09-083-116-5	Sequence 5, Appli
13	424.5	6.2	2035	US-08-479-537A-2	Sequence 2, Appli
14	424.5	6.2	2035	US-09-083-116-2	Sequence 2, Appli
15	419.5	6.2	2476	US-08-276-967-2	Sequence 2, Appli
16	419	6.1	1185	US-09-041-886-23	Sequence 23, Appli
17	417	6.1	829	US-08-642-255-132	Sequence 132, Appl
18	417	6.1	829	US-08-397-633A-33	Sequence 53, Appli
19	417	6.1	829	US-08-175-155-68	Sequence 68, Appli
20	417	6.1	837	US-08-477-509B-103	Sequence 103, App
21	417	6.1	837	US-08-642-255-101	Sequence 101, App
22	417	6.1	837	US-08-707-237A-75	Sequence 75, Appli
23	417	6.1	837	US-08-482-085B-103	Sequence 103, Appl
24	417	6.1	837	US-08-397-633A-50	Sequence 50, Appli
25	413.5	6.1	907	US-08-783-774-2	Sequence 2, Appli
26	413.5	6.1	907	PCT-US95-04611A-19	Sequence 19, Appli
27	404.5	5.9	408	US-07-609-716-65	Sequence 65, Appli

28	404.5	5.9	408	US-08-475-411A-65	Sequence 65, Appli
29	404.5	5.9	408	US-08-478-029A-65	Sequence 65, Appli
30	398.5	5.8	682	US-08-642-255-126	Sequence 126, App
31	398.5	5.8	682	US-08-397-633A-36	Sequence 36, Appli
32	381	5.6	1848	US-08-296-791-6	Sequence 6, Appli
33	381	5.6	1848	PCT-US95-10661A-6	Sequence 6, Appli
34	376	5.5	960	US-09-219-849-6	Sequence 6, Appli
35	375.5	5.5	1231	US-08-904-263A-4	Sequence 84, Appli
36	374.5	5.5	1537	US-08-325-267A-2	Sequence 2, Appli
37	371	5.4	761	US-08-707-237A-84	Sequence 114, App
38	371	5.4	762	US-08-642-255-114	Sequence 20, Appli
39	371	5.4	762	US-08-397-633A-26	Sequence 62, Appli
40	370.5	5.4	1064	US-08-642-255-62	Sequence 120, Appl
41	368	5.4	762	US-08-642-255-120	Sequence 31, Appli
42	368	5.4	762	US-08-397-633A-31	Sequence 28, Appli
43	367.5	5.4	1187	US-08-320-559-28	Sequence 28, Appli
44	367.5	5.4	1187	US-08-545-860D-28	Sequence 28, Appli
45	367.5	5.4	1187	PCT-US94-04496-28	Sequence 28, Appli

ALIGNMENTS

```
RESULT 1
; Sequence 4, Application US/09103429A
; Patent No. 6187558
; GENERAL INFORMATION:
; APPLICANT: Granados, Robert R
; APPLICANT: Wang, Ping
; TITLE OF INVENTION: A No. 6187558el Invertebrate Intestinal Mucin
; TITLE OF INVENTION: cDNA and Related Products and Methods
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Piniel & Michaels, P.C.
; STREET: 118 No. 6187558th Tlaga
; CITY: Ithaca
; STATE: NY
; COUNTRY: USA
; ZIP: 14850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/103,429A
; FILING DATE: 24-JUN-1998
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Michaels, Christopher A
; REGISTRATION NUMBER: 34,390
; REFERENCE/DOCKET NUMBER: BTI-39
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (607) 256-2000
; TELEFAX: (607) 256-3628
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 805 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Trichoplusia ni
; TISSUE TYPE: peritrophic membrane
; US-09-103-429A-4
; Query Match 7.7%; Score 522; DB 4; Length 805;
; Best Local Similarity 27.8%; Pred. No. 6.5e-27;
```



SEQUENCE 1970 AA: 217205 MW: 6876FC25692A657E CRC64;

Query Match 7.4%; Score 503.5; DB 1; Length 1970;
Best Local Similarity 33.3%; Pred. No. 6.8e-14;
Matches 181; Conservative 101; Mismatches 167; Indels 95; Gaps 47;

```

OY 216 PPTTTPKEPASTTPKE-PPTTTPKSAPTTTPKEPAPTTTTPKSAPTTTPKE-----PAPTTPKE 269
DB 1507 PSPMGISIPAMTPMNGATATATGAMSSVSGMTPGAAGSPSAASDAGSPGSPAWMS 1566
OY 270 PAPTTPKEPAPTTPKEPAPTTPKSAPTTTPKEPAPTTPKKAAPTTPKEP--APTTPKEPTP 327
DB 1567 PTPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 1624
OY 328 TTPKEPAPTTPKEPAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPK 387
DB 1625 TSPSPY-SPTSPNYSPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY 1671
OY 388 PAPTTPKSAPTTTPKEPAPTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTP 447
DB 1672 -SPTSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 1725
OY 448 KEPAPTTPKEPAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKA 507
DB 1726 SY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSP 1777
OY 508 PTPPELAPTPPEPPTTPPEPAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKA 565
DB 1778 PTSP-NYSPSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSP 1828
OY 566 TPKEAPTTPKGTATPTTKEPAPTTPK-KAPAKELAPTTTKEPTSTSDKAPATTPKGT 624
DB 1829 SP-SYSPSPSKYT-----PTSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 1875
OY 625 PTPPEPAPTTPKEPAPTTPKGTATPTTKEPAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKA 683
DB 1876 PTSPSY-SPTTPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSP 1922
OY 684 PAPTTPKETAAPTTPKEP--APTTPKKAAPTTPPETTPPTTSEVSTTTTKEPTTIKSPDE 741
DB 1923 YSPSPY-TYSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 1965
OY 742 STPE 745
DB 1966 SDEE 1969

```

RESULT 14
RPTL_MOUSE STANDARD: PRT: 1970 AA.
ID RPTL_MOUSE
AC P08775;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DNA-DIRECTED RNA POLYMERASE II LARGEST SUBUNIT (EC 2.7.7.6) (RPL1).
GN POLRA OR RPO2-1 OR RPII25.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID-10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-87280135; PubMed-3038894;
RA Ahearn J.M. Jr., Bartolomei M.S., West M.L., Cisek L.J., Corden J.L.;
RT Cloning and sequence analysis of the mouse genomic locus encoding
RT the largest subunit of RNA polymerase II.;
RL J. Biol. Chem. 262:10695-10705(1987).
[2]
RP SEQUENCE OF 1587-1970 FROM N.A.
RX MEDLINE-86068017; PubMed-299785;
RA Corden J.L., Cadena D.L., Ahearn J.M. Jr., Dahmus M.E.;
RT A unique structure at the carboxyl terminus of the largest subunit

RT of eukaryotic RNA polymerase II.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:7934-7938(1985).
RN [3]
RP REVISIONS, AND PRESENCE OF AN ADDITIONAL EXON.
RX MEDLINE-92178992; PubMed-1542581;
RA Winzerlth M., Acker J., Vicaire S., Vigneron M., Keding C.;
RT "Complete sequence of the human RNA polymerase II largest subunit.";
RL Nucleic Acids Res. 20:910-910(1992).
CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC -1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N PYROPHOSPHATE +
CC RNA(N).
CC -1- SUBUNIT: RNA POLYMERASE II CONSISTS OF 12 DIFFERENT SUBUNITS.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- PTM: THE TANDEM 7 RESIDUES REPEATS CAN BE HIGHLY PHOSPHORYLATED.
CC -1- THE PHOSPHORYLATION ACTIVATES POL2.
CC -1- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE
CC FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA
CC PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE
CC III FOR 5S AND TRNA GENES.
CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
CC or send an email to license@isb-sdb.ch).
CC -----
DR EMBL: M12130; AAA40071.1; -;
DR EMBL: M14101; AAA40071.1; JOINED.
DR PIR: A28490; A28490.
DR MGD: MGI:98086; Rpo2-1.
DR InterPro: IPR000684; Rpo2-1.
DR InterPro: IPR000722; RNA_POL_A.
DR InterPro: IPR002879; RNA_POL_A2.
DR Pfam: PF00623; RNA_POL_A; 1.
DR Pfam: PF01854; RNA_POL_A2; 1.
DR PROSITE: PS00115; RNA_POL_II_REPEAT; 42.
DR TRANSFERASE: DNA-directed RNA polymerase; transcription; zinc; Repeat;
KW DNA-binding; Nuclear protein; Phosphorylation; zinc-finger.
KW ZN-FING 71
FT DOMAIN 1590 1958 CARBOXYL-TERMINAL 7-RESIDUE REPEATS.
FT CONFLICT 1498 1498 P->R (IN REF. 1 AND 2).
FT CONFLICT 1499 1536 MISSING (IN REF. 1 AND 2).
FT SEQUENCE 1970 AA; 217175 MW; 7D76F38FD92A657E CRC64;

Query Match 7.3%; Score 498.5; DB 1; Length 1970;
Best Local Similarity 33.1%; Pred. No. 1.1e-13;
Matches 180; Conservative 101; Mismatches 168; Indels 95; Gaps 47;

```

OY 216 PPTTTPKEPASTTPKE-PPTTTPKSAPTTTPKEPAPTTPKSAPTTTPKE-----PAPTTPKE 269
DB 1507 PSPMGISIPAMTPMNGATATATGAMSSVSGMTPGAAGSPSAASDAGSPGSPAWMS 1566
OY 270 PAPTTPKEPAPTTPKEPAPTTPKSAPTTTPKEPAPTTPKKAAPTTPKEP--APTTPKEPTP 327
DB 1567 PTPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 1624
OY 328 TTPKEPAPTTPKEPAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPK 387
DB 1625 TSPSPY-SPTSPNYSPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY 1671
OY 388 PAPTTPKSAPTTTPKEPAPTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTP 447
DB 1672 -SPTSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 1725
OY 448 KEPAPTTPKEPAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKA 507
DB 1726 SY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSP 1777

```

FT DISULEID 374 391 BY SIMILARITY.
 FT DISULEID 526 552 BY SIMILARITY.
 FT DISULEID 536 551 BY SIMILARITY.
 FT DISULEID 546 563 BY SIMILARITY.
 FT DISULEID 573 599 BY SIMILARITY.
 FT DISULEID 583 598 BY SIMILARITY.
 FT DISULEID 593 610 BY SIMILARITY.
 FT DISULEID 621 647 BY SIMILARITY.
 FT DISULEID 631 646 BY SIMILARITY.
 FT DISULEID 641 658 BY SIMILARITY.
 FT VARIANT 276 276 K -> E.
 FT VARIANT 354 354 C -> R.
 FT VARIANT 415 415 T -> A.
 SO SEQUENCE 662 AA: 67774 MW: F085277F1ED2FD40 CRC64;

Query Match 7.5%; Score 509.5; DB 1; Length 662;
 Best Local Similarity 28.1%; Pred. No. 1.6e-14;
 Matches 223; Conservative 48; Mismatches 276; Indels 247; Gaps 30;

QY 157 TTKKOTSDCKEKTAKETOSTIEKTSADLAPTSKVLAKPTPKAETTTKGPALTTPKEP 216
 DB 3 TTAATAAAGCKDTAAAECSAAAEKTA-----AGEVSAPV--AAVATGEDATT----- 51
 QY 217 TPTTPKEPASTPKEPTPTTIKSAPTTKEPAPTTSKAPTTPKEPAPTTPKEPAPTTPK 276
 DB 52 -----AAATTAATTTAAAGAPATTTPATTAAGKAPTAAATAPTTAAAGAPTAT 103
 QY 277 EPAPTTPKAPAPTTPKS--APTTPKEPAPTTPKAPAPTTPKEPAPTTPKEPAPTTPK 334
 DB 104 GKAPATAAAPVPTTAASKAPTAAATHTAAAPTAATAASAKKKESTSSSEEHCH 163
 QY 335 --PTTPKEAPT--TPKEAPTAPKAPAPTTPK-----EPAPTTPKEPAPTTPKE 379
 DB 164 VAPSKREKSGSGITRKO-----CKKKNCCEPDKGHGIGHCEHKKKSHHEHTTTTK- 218
 QY 380 PPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 439
 DB 219 -----APTPTQIATTTT--TPPTT-----TTTTKAPTTP----- 245
 QY 440 KKPAPTTPKEPAPTTPKEPAPTTPKAPAPAKKEPAPTTPKEPAPTTPKLTPTTEKLA 499
 DB 246 -----TTTKATPTTP-----TTTKATTTT-----TPPT----- 270
 QY 500 PTPPEKAPATTPPEELAPTPPEEPPTTPPEEPAPTTPKAAANTKEPAPTTPKEPAPTTP 559
 DB 271 -----TTTTKATTTPTTTTTTP----- 289
 QY 560 KEPAPTTPKEPAPTTPKGTATPTTLKEPAPTTPKAPAKELAPTTTKEPTSTSDKAPTTP 619
 DB 290 -----TTTTKATTTTTTSGCKMEPSK-----REDCGSGITTESOCR 328
 QY 620 PKG-----TAPPT-----PKEPAPTTPKEPAPTTPKGNAPTTLKEPAPTTPKAPAPELA 669
 DB 329 TKGCFDSSIQTKWCFYTLQVADCKVEFSQAVDCRGIT-----ADOCROKNCCEPSS 384
 QY 670 PTTTKGPTSTSDKAPAPTTPKEPAPTTPKAPAPTTPKAPAPTTPPTTPPTSEVSTPTTP 729
 DB 385 ISGKWCFFYSOVAA--TKTTTTPTTTTTPTTTTTTKATTTTP-----TTTTTTTTTT 436
 QY 730 KEPTTIHKSPEDESPELSAEPPTKALNSPKREVPTTKPAATKPEMTTTAANDKTTTERD 789
 DB 437 TTTT-----TTKA-----TTTTPTTTTTTTTTTT-----KAT----- 463
 QY 790 LRTTPEPTTAAPKMTKETATTEKTESKTTATTTQVSTTTODTTPPEKLTTLTKTTTLAP 849
 DB 464 -TTTTPTTTTTTP-----TTTTTKAT-----TTTTTTTTTTTTTTTTTTTTTTTT 510
 QY 850 KVTITKTTTITTEIMNKPEETAPKORATSKATTPKQKPTAAPKPKSTKPKP-----T 905
 DB 511 TTTTAKTTTTTSGCKME-----PSKADCGPGITGESOCRSKGCCEPSSIQTKWCFYS 566
 QY 906 MPVRKRPKPTTPTR 919

DB 567 LPQVADCKVAPSSR 580
 RESULT 13
 ID REP1_HUMAN STANDARD; PRT; 1970 AA.
 AC P24928;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE DNA-DIRECTED RNA POLYMERASE II LARGEST SUBUNIT (EC 2.7.7.6) (RPB1).
 GN POLR2A.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92178992; PubMed=1542581;
 RA Wintzerith M., Acker J., Vicaire S., Vigneron M., Keding C.;
 "Complete sequence of the human RNA polymerase II largest subunit.";
 Nucleic Acids Res. 20:910-910(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95347616; PubMed=7622068;
 RA Mita K., Tsuji H., Moriyama M., Takahashi E., Nenoi M.,
 Ichimura S., Yamauchi M., Hongo E., Hayashi A.;
 "The human gene encoding the largest subunit of RNA polymerase II.";
 Gene 159:285-286(1995).
 CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
 OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
 SUBSTRATES.
 CC -1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE - N PYROPHOSPHATE +
 RNA(N).
 CC -1- SUBUNIT: RNA POLYMERASE II CONSISTS OF 12 DIFFERENT SUBUNITS.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- PTM: THE TANDEN 7 RESIDUES REPEATS CAN BE HIGHLY PHOSPHORYLATED.
 CC THE PHOSPHORYLATION ACTIVATES POL2.
 CC -1- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE
 FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA
 PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE
 III FOR 5S AND TRNA GENES.
 CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch)
 CC -----
 DR EMBL: X63564; CAA45125.1;
 DR EMBL: X74874; CAA52862.1;
 DR EMBL: X74873; CAA52862.1; JOINED.
 DR EMBL: X74872; CAA52862.1; JOINED.
 DR EMBL: X74871; CAA52862.1; JOINED.
 DR EMBL: X74870; CAA52862.1; JOINED.
 DR PIR: S21054; S21054.
 DR MIM: 180660;
 DR InterPro: IPR000684; RNA_polII_repeat.
 DR InterPro: IPR000722; RNA_pol_A.
 DR InterPro: IPR002879; RNA_pol_A2.
 DR Pfam: PF00623; RNA_pol_A_1.
 DR Pfam: PF01854; RNA_pol_A2_1.
 DR PROSITE: PS00115; RNA_POL_II_REPEAT; 43.
 KW Transferrase; DNA-directed RNA polymerase; Transcription; Zinc; Repeat;
 KW DNA-binding; Nuclear protein; Phosphorylation; Zinc-finger.
 FT ZN_FING 71 87
 FT DOMAIN 1590 1958 C2H2-type (POTENTIAL).
 FT CONFLICT 1067 1067 W -> L (IN REF. 2).
 FT CONFLICT 1449 1449 D -> Y (IN REF. 2).

```

Db 186 RKQSYTPYKPKATYPPTTKKITYP-----PTYKKRPSY-----PYKKPTTYPTTKPKKI 236
QY 309 PAPTKEKPAPTT-----KPEPIPTT--KPEAPTTKEPAPTTTKEKAPL-APKKAPPTP 361
Db 237 SYSSIVKPKKASYSYSKSKTYPPTTKPKISYPTTKPKHSYP-----PTYKKVITYPTYP 292
QY 362 KEKAPPTKKEPAPTTTKEPPTTKKEPAPTTTKSAPL-----TTKEPAPTTTKSAPTTPK 416
Db 293 K-KPKSPYPTTKKITYPTT-PYKPSYPTTKYOKPSYPTTKSKSYPTTKSKSYPTTKSKSYPTTK 350
QY 417 EPSPPTTKEBA-PPTKEKAPTTTTPKRP-APL-TPK-----EPAPTTTKEPAPTTTTPK 466
Db 351 TYKKKITYPTTKPKKPSYPSYKPKKITYSPPTTKKITYPTTKPKKPSYPSYKPKKITYPTTK 410
QY 467 APAPKAPAPTTTKEKAP-----TTTPKKLPTTPEKLAPTTPEKAPTTTPEELAPTTPEE 521
Db 411 T-YKPKISYPTTKPKKASYSYSKSKTYPPTTKPKISYPTTKPKKPSYPTTKPKKITYPTTK 469
QY 522 PT-PPTPEEPAPTTPKAAAPNTPK-EPAPTTTKEKAPPTP-----KEKAPTT--PKETAP 572
Db 470 PTYKPKPSYPTTKKITYPTTKPKKPSYPTTKYOKPSYPTTKSKSYPTTKSKSYPTTKSKSYPTTK 529
QY 573 TT--PKGTAPTTKEKAPTTTTPKRPAPKELAPTTTKEPSTSDKAPL-TPKGTAPTTPK 629
Db 530 PTYKPKKITYPTTK-----PKPSYPSYKPKKITYPTTKPKKIKITYPTTKPKKASYPPTTK 583
QY 630 EPAPTTTKEKAPTTTTPKGTAPTTTKEKAPPTP-----KKRPAPKELAPTTTTPKGTSTSDKPA 685
Db 564 -----PKITYPTTK-PKPSYPTTKPKKPSYPTTKPKKPSYPTTKSKSYPTTKSKSYPTTKSKSYPTTK 628
QY 666 PTT--PKETAPTTTKEKAPTTPK--KRAPTTTPEP-----PPTTTSVSPTTTPKEPTT 734
Db 629 PPTTKKITYPTTK-PKPSYPSYKPKKITYPTTKPKKPSYPSYKPKKITYPTTKPKKPSYPSYKPKK 687
QY 735 IHKSPDESTPELSAPPTPKALENSPEKPGVPTTKTPATKPEMTTANKT-----TERDL 790
Db 688 Y---PSPKPKKITYPTTK-----PKISYPTTK-----TKPSYPSYKPKKPSYPTTKPK 735
QY 791 RTTPETTTAAPKMTKEATTEKTESKITATTTTOVTTTODITTPKKITTLKTTLAPK 850
Db 736 ISYPTTKAKP-----SYPTTKPKKPSYASYSKPKIRYPTTKPKKPSYASYSK-----PK 785
QY 851 VTTTKKITTITIMKPEETAKPKDRATNSKATTPKPO---KPTKAPKRP--TSTKPKP-T 905
Db 786 I-----RYPTTKPKP-----SYASYSKPKIRYPTTKPKKPSYASYSKPKKIT 827
QY 906 MPVRKPKPTT--PTPKMTSTMPKELNPTSRILAEMLQTTTPNQ 947
Db 828 YPTTKPKKISYPTTKKITYPTTKPKKISYPTTKPKKISYPTTKPKKISYPTTKPKKISYPTTKPK 871

```

```

CC -1- FUNCTION: COULD BE INVOLVED IN DEFENSE AGAINST MICROBIAL
CC INFECTIONS. PROTECTS THE EPITHELIA FROM EXTERNAL ENVIRONMENT.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- ALTERNATIVE PRODUCTS: A NUMBER OF DIFFERENT FORMS OF THE PROTEIN
CC MAY BE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: SKIN.
CC -1- PTM: EXTENSIVELY O-GLYCOSYLATED.
CC -1- SIMILARITY: CONTAINS 6 P-TYPE (TREPOLL) DOMAINS.
CC THE SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION
CC BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION -
CC THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS
CC USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY
CC MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL
CC ENTITIES REQUIRES A LICENSE AGREEMENT (SEE http://www.isb.ch/announce/
CC OR SEND AN EMAIL TO license@isb.slb.ch).
CC -----
CC EMBL: L02115; AAA74725.1; -.
CC PIR: A45155; A45155.
CC HSSP: P04002; 1WFA.
CC InterPro: IPR000519; P_trepoll, 6.
CC Pfam: PF00088; trefoll; 6.
CC SMART: SM00018; P_6.
CC PROSITE: PS00025; P_TREPOLL; 6.
CC Repeat: Amphibian skin; Glycoprotein; Alternative splicing.
CC NON_TER 1 144
CC DOMAIN 1 144
CC FT 81 88
CC REPEAT 81 88
CC FT 89 96
CC REPEAT 89 96
CC FT 97 104
CC REPEAT 97 104
CC FT 105 112
CC REPEAT 105 112
CC FT 113 120
CC REPEAT 113 120
CC FT 121 128
CC REPEAT 121 128
CC FT 129 136
CC REPEAT 129 136
CC FT 137 144
CC REPEAT 137 144
CC FT 161 202
CC REPEAT 161 202
CC FT 218 301
CC REPEAT 218 301
CC FT 218 224
CC REPEAT 218 224
CC FT 225 239
CC REPEAT 225 239
CC FT 240 249
CC REPEAT 240 249
CC FT 250 259
CC REPEAT 250 259
CC FT 260 275
CC REPEAT 260 275
CC FT 276 287
CC REPEAT 276 287
CC FT 288 294
CC REPEAT 288 294
CC FT 295 301
CC REPEAT 295 301
CC FT 306 347
CC REPEAT 306 347
CC FT 353 394
CC REPEAT 353 394
CC FT 402 522
CC REPEAT 402 522
CC FT 402 411
CC REPEAT 402 411
CC FT 412 419
CC REPEAT 412 419
CC FT 420 431
CC REPEAT 420 431
CC FT 432 443
CC REPEAT 432 443
CC FT 444 453
CC REPEAT 444 453
CC FT 454 460
CC REPEAT 454 460
CC FT 461 472
CC REPEAT 461 472
CC FT 473 479
CC REPEAT 473 479
CC FT 480 491
CC REPEAT 480 491
CC FT 492 498
CC REPEAT 492 498
CC FT 499 515
CC REPEAT 499 515
CC FT 516 522
CC REPEAT 516 522
CC FT 525 566
CC REPEAT 525 566
CC FT 572 613
CC REPEAT 572 613
CC FT 620 661
CC REPEAT 620 661
CC FT 162 188
CC DISULFID 162 188
CC FT 172 187
CC DISULFID 172 187
CC FT 182 199
CC DISULFID 182 199
CC FT 307 333
CC DISULFID 307 333
CC FT 317 332
CC DISULFID 317 332
CC FT 327 344
CC DISULFID 327 344
CC FT 354 380
CC DISULFID 354 380
CC FT 364 379
CC DISULFID 364 379

```

```

3-1.
3-2.
3-3.
3-4.
3-5.
3-6.
3-7.
3-8.
3-9.
3-10.
3-11.
3-12.
P-TYPE 4.
P-TYPE 5.
P-TYPE 6.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
12 X APPROXIMATE TANDEM REPEATS,
THR-RICH.

```



```

OY      662 KPAKEALPTTTKGSTSTSDKAPATTPK-----ETAPTTPKCAPPTPKKPAPTPEPTEP 717
       | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1034 TP-----ADSSAHGTPSTPADSSASHSP-PPVDSASPSSTPADSS-----AGCTSTPA 1084
OY      718 PTISEVSTPTTTRKPTTHHKSPPDESTP-EISAEPTEKALENSPKPEGVPPTKTPTAATKE 776
       | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1085 DSSAH-STPSTPAD-SSAHGP--STPADSSASHSTP-----STPADSSAN 1125
OY      777 MT 778
Db      1126 GT 1127

RESULT 10
CPN_DROME CPN_DROME STANDARD: PRT: 865 AA.
AC 002910;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE CALPHOTIN.
GN CPN OR CAP.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Phlebotomidae; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CANTON-S;
RC MEDLINE=93165729; PubMed=8094559;
RA Martin J.H., Benzer S., Rudnicka M., Miller C.A.;
RT "Calphotin: a Drosophila photoreceptor cell calcium-binding protein.";
RL proc. Natl. Acad. Sci. U.S.A. 90:1531-1535(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CANTON-S;
RC MEDLINE=93165730; PubMed=8434015;
RA Ballinger D.G., Xue N., Harshman K.D.;
RT "A Drosophila photoreceptor cell-specific protein, calphotin, binds
calcium and contains a leucine zipper.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:1536-1540(1993).
CC -1- FUNCTION: MIGHT FUNCTION OF A CALCIUM-SQUOSTERING "SPONGE" TO
REGULATE THE AMOUNT OF FREE CYTOPLASMIC CALCIUM. IT BINDS 0.3 MOL
OF CA+2 PER MOL OF PROTEIN.
CC -1- SUBUNIT: HOMODIMER (PROBABLE).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC; HYPODENSE COMPARTMENT.
CC -1- TISSUE SPECIFICITY: SOMA AND AXONS OF PHOTORECEPTOR CELLS OF
COMPOUND EYES AND OCCELLI.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED EARLY IN PHOTORECEPTOR CELL
DEVELOPMENT.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isd-sib.ch/announce/
or send an email to licenses@isb-sib.ch).
CC -----
DR EMBL: L02111; AAA28405.1; -
DR EMBL: L05080; AAA28420.1; -
DR PIR: A47282; A47282.
DR FlyBase: FBgn0010218; Cpn.
KW Calcium-binding.
FT CONFILCT 36 36 A -> AVAPAVVA (IN REF. 2).
FT CONFILCT 43 43 I -> T (IN REF. 2).
FT CONFILCT 64 64 I -> V (IN REF. 2).
FT CONFILCT 76 76 T -> A (IN REF. 2).
FT CONFILCT 100 100 P -> PP (IN REF. 2).
FT CONFILCT 127 127 VO -> AP (IN REF. 2).
FT CONFILCT 154 154 I -> V (IN REF. 2).
FT CONFILCT 160 160 S -> T (IN REF. 2).

```

TO 716 AND IS SHORTER DUE TO FRAMESHIFTS.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: M24496; AAA39813.1; -
 DR EMBL: M23349; AAA39813.1; JOINED.
 DR EMBL: M24494; AAA39813.1; JOINED.
 DR EMBL: M24495; AAA39813.1; JOINED.
 DR EMBL: M35131; AAA39809.1; ALT_FRAME.
 DR EMBL: M31012; CAA83229.1; -
 DR PIR: J10368; QFMH.
 DR PIR: A43778; A43778.
 DR MGI: MGI:97309; NFH.
 DR InterPro: IPR001664; IF.
 DR Pfam: PF00038; Filament; 1.
 DR PROSITE: PS00226; IF; 1.
 KW Intermediate filament; Coiled coil; Neurone; Phosphorylation;
 Repeat.
 FT DOMAIN 1 97 HEAD.
 FT 98 408 ROD.
 FT 409 1087 TAIL.
 FT 436 517 GLU-RICH (ACIDIC).
 FT 519 886 50 X 6 AA TANDEM REPEATS OF K-S-P-A-E-A.
 FT 887 1087 GLU/LYS-RICH.
 FT 98 129 COIL 1A.
 FT 130 141 LINKER 1.
 FT 142 239 COIL 1B.
 FT 240 261 LINKER 12.
 FT 262 283 COIL 2A.
 FT 284 287 LINKER 2.
 FT 288 408 COIL 2B.
 FT 133 133 K -> QA (IN REF. 2 AND 3).
 FT 133 133 A -> AR (IN REF. 2 AND 3).
 FT 199 199 S -> T (IN REF. 2 AND 3).
 FT 281 281 L -> G (IN REF. 2 AND 3).
 FT 492 492 P -> PREAKSP (IN REF. 3).
 FT 551 551 MISSING (IN REF. 3).
 FT 689 712 G -> A (IN REF. 3).
 FT 714 714 V -> M (IN REF. 2 AND 3).
 FT 814 814 T -> N (IN REF. 2 AND 3).
 FT 843 843
 SO SEQUENCE 1087 AA; 116612 MW; 57BAC76A38EDICB9 CRC64;

Query Match 7.8%; Score 533; DB 1; Length 1087;
 Best Local Similarity 27.5%; Pred. No. 2,7e-15;
 Matches 216; Conservative 81; Mismatches 330; Indels 158; Gaps 38;

OY 86 DEASGIDNGEFKYT-----TPDSTTQHNNKVTSPRIITAKPINRPSLPNSDTSKET 140
 DB 404 EECIGGSPSPSLTEGLPIPIPSIST--HIKVKSEEMIKVVE-----KSEKET 449
 OY 141 SLTVNKETVETKETTNTKOTSDGKEKTTSAKETOSIKTSKADLAFTSKVLAPTRK 200
 DB 450 VIVGGEERIRVTEGVTEEDKKGOGEGEAEKEKEEELAAATSPAAEAASPEKE 509
 OY 201 AETTKAPALTTRKEPTPTTPKPEASTTPKEPTPTTIKSAPTTPKPEAPTTPKSAPTTPK 260
 DB 510 TKSIVKEEAKSPGKSPGAKSPA---EAKSPGKKS--PGGAKSPGKSPGAKSPAEKSPA 564
 OY 261 EP-APTTKPEAPTTTPKPEAPTTPKPE---APTTPKSAPTTPKPEA---PTTPKKA- 310
 DB 565 EKPSPAKAKSPA---EKSPA--TVKSPGKAKSPSEAKS--PAEAKSPAEEKSPAEEKSPA 619
 OY 311 ---PTTPKPEA---PTTPKPE---TPTTPKPEAPTTPKPEAPTTPKPEA---PTPAK 355
 DB 620 AKSPAEEKSPAEEKSPATYKSPGKSPGAKSPA---EAKSPAEEKSPAEEKSPAEEKSPA 676

OY 356 P-----APTTKPEAPTTTPKPEAPTTPKPEAPTTPKPEA-----PTTPKSAPTTPKPEAPT 407
 DB 677 PGEAKSPAEPKSPAPAA--KSPA-----EYKSPAEEKSPAEEKSPGKAS--PAVNSPAEAK 729
 OY 408 TKSAPTTPKE--PSPTTPKEA-----PTTPKPEAPTTPKPAAPTTPKPEAPTTPKPEAPT 462
 DB 730 SPAAKSPGKAKSPGKSPGAKSPAEEKSPAEEKSPAEEKSPAEEKSPAEEKSPAEEKSPA 789
 OY 463 TKKPAPTAPKPEAPTTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPE 522
 DB 790 AKSPKPEKIDIKPPAEKSPKAEK--KSPVKEGAKPPKAPPLDVKSPGAGTPOVEATVPTDI 848
 OY 523 TPTTPKE--PAPTTPKAAAPNTPKPEAPTTP-----KPEAPPTTP-----KPEAPTTPKET 570
 DB 849 RP--PEOVKSPAKKAKSP--KEEAKTSEKAVKKEGVNSPVKEVAKKEPKKVEEK 904
 OY 571 APTTPKGAPTTPKPEAPTTPKPAPELAPTTTKEPTSTSD--KPAPTTPKGAAPTTP 628
 DB 905 TLPTTPKEAKSKKDEAPKPAKPKVEKEKTEPKPKDSTAEEKKEAGEKKKAVASEE 964
 OY 629 KEAPPTTPKPEAPTTPKGAAPTTPKPEAPTTPKPAPELAPTTTKEPTSTSDOKAPPT 688
 DB 965 ETPAKLGVKEEA--KPKKETETTEADTKAKEPS---KPTETEK----- 1006
 OY 689 PKETAPTTKPEAPTTPKPAAPTTPPETPTTSEVSTPTTPKPTTIKSPDSTPELSA 748
 DB 1007 -----KKEEMPAEK-----KDKTEKTESRKEEK----- 1034
 OY 749 EPTPKALENSPEKGVPTTPKPAATKPEMTTAKDKTTEEDLTTPETTTAAKMKTEA 808
 DB 1035 ---PK-NEAVKEDDKSLSEP--SKPTKAKSSSTIDKESOPSE-----KTTEKRA 1082
 OY 809 TTEK 813
 DB 1083 TKGER 1087

RESULT 9
 TCNA_TRYCR
 ID TCNA_TRYCR STANDARD; PRT; 1162 AA.
 AC P23253;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE SIALIDASE (EC 3.2.1.18) (NEURAMINIDASE) (NA) (MAJOR SURFACE ANTIGEN).
 GN TCNA.
 OS Trypanosoma cruzi.
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
 OX Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
 RN NCBI_TaxID=5693;
 RP SEQUENCE FROM N.A.
 RC STRAIN-SILVIO X-10/4;
 RX MEDLINE=91277609; PubMed=1711561;
 RA Pereira M.E.A., Mejia J.S., Ortega-Barria E., Matzilevich D.,
 RA Prioli R.P.;
 RT "The Trypanosoma cruzi neuraminidase contains sequences similar to
 RT bacterial neuraminidases, YWRD repeats of the low density lipoprotein
 RT receptor, and type III modules of fibronectin.";
 RL J. Exp. Med. 174:1179-191(1991).
 RN [2]
 RP SUBCELLULAR LOCATION.
 RX MEDLINE=91376547; PubMed=1896773;
 RA Prioli R.P., Mejia J.S., Aji T., Alkawa M., Pereira M.E.A.;
 RT Trypanosoma cruzi: localization of neuraminidase on the surface of
 RT trypanomastigotes.";
 RL Trop. Med. Parasitol. 42:146-150(1991).
 CC -!- FUNCTION: DEVELOPMENTALLY REGULATED NEURAMINIDASE IMPLICATED IN
 CC PARASITE INVASION OF CELLS.
 CC -!- CATALYTIC ACTIVITY: HYDROLYSIS OF 2,3'- AND 2,8-GLYCOSIDIC
 CC LINKAGES JOINING TERMINAL NON-REDUCING N- OR O-ACYLENEURAMINYL
 CC RESIDUES TO GALACTOSE, N-ACETYLHEXOSAMINE, OR N- OR O-ACETLATED
 CC NEURAMINYL RESIDUES IN OLIGOSACCHARIDES, GLYCOPROTEINS,
 CC GLYCOLIPIDS OR COLOMINIC ACID.

01-MAR-1992 (rel. 21, Last annotation update)
 DE EXTENSIN PRECURSOR (CELL WALL HYDROXYPROLINE-RICH GLYCOPROTEIN).
 GN HRGPPT3.
 OS Nicotiana tabacum (Common tobacco);
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 CC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
 NCBI_TaxID=4097;
 OX (1)
 RN SEQUENCE FROM N.A.
 RC STRAIN=CV. XANTHI; TISSUE=Leaf;
 RX MEDLINE=90128263; PubMed=2612909;
 RA Keller B., Lamb C.J.;
 RT "Specific expression of a novel cell wall hydroxyproline-rich
 glycoprotein gene in lateral root initiation."
 RL Genes Dev. 3:1639-1646(1989).
 CC -1- FUNCTION: HAS A SPECIALIZED STRUCTURAL FUNCTION, POSSIBLY IN
 THE MECHANICAL PENETRATION OF THE CORTEX AND EPIDERMIS OF THE
 MAIN ROOT.
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX.
 CC -1- PTM: EXTENSINS CONTAIN A CHARACTERISTIC REPEAT OF THE PENTAPEPTIDE
 SER-PRO(4). THE PROLINE RESIDUE IS HYDROXYLATED AND THEN
 GLYCOSYLATED.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X13885; CAA32090.1; -.
 DR PIR: S06733; S06733.
 KW Repeat; Cell wall; Glycoprotein; Signal; Structural protein;
 KW Hydroxylation.
 FT SIGNAL 1 ?
 FT CHAIN 70
 FT REPEAT 73
 FT REPEAT 148 151 H-A-P-P.
 FT DOMAIN 229 242 2 X 7 AA TANDEM REPEATS OF T-H-R-H-A-P-P.
 FT REPEAT 229 235 1.
 FT REPEAT 236 242 2.
 FT REPEAT 205 620 CONTAINS THE SER-PRO(4) REPEATS.
 FT DOMAIN 499 600 3 X APPROXIMATE TANDEM REPEATS.
 FT DOMAIN
 SQ SEQUENCE 620 AA: 65406 MW: 641DD2278AB28524 CXC64;

Query Match 8.1%; Score 551; DB 1; Length 620;
 Best Local Similarity 27.2%; Pred. No. 3.1e-16;
 Matches 170; Conservative 64; Mismatches 304; Indels 88; Gaps 24;

OY 201 AETTKR-GPALTTP--KEPTPT-----PKEPASTTPKE--PTPTTKSAPTTKKEAP 249
 DB 24 AEATQYGYGLDPPVYSGPPSSIGLSPSAPTTTPPSGHWSP--NHAPRNAVPP 80
 OY 250 TTTKSAPTTKPE-----APTTEKAPPT--TPKEPATTKEAPTTKSAPTTKPE 300
 DB 81 SHCHLPSSGGPPRHGHLPPSGENFPSPVISPNSHPPSGAPRRSGRCHLDSHGCR 140
 OY 301 PAPTTPKKAAPT--TPKEPAPTTPKEPTTPPTTKKEAPTTK--EAPTTKKEAPAPA 355
 DB 141 P-----PSPSHGAPPSGCHTPPRGQHPSPSHRPSRCHNGHPPRYAOPRPPIYS 193
 OY 356 PAPTTPKKAAPTTPKEPAPTTPKEPTTPK--EAPTTKSAPTTKKEAPTTK-- 410
 DB 194 PSHOV--QRPPT--YSPRPTHQPTPSPSRCHQOPRTHRAAPTTHKHAAPTTHQPSPL 249
 OY 411 ---APTTPKPSPTTKKEAPTTKKEAPTTKKAAPTTPKKAAPT--TPKEPAPT 464
 DB 250 RHLPPSRKRPQRPPTYSPPRAVAQSPQSPPTYSPPPTYSPPPSPIYSPPPAPSPSP 309
 OY 465 KPAPT--AKKEPAPTTPKETAPTTKKL-TPIT-----PEKLATTPKKAAPTTPPEL 514

DB 310 PPTPTTPSSPPPAASPPPTYSPPPTLLPLSSPIYSPPPVYSPPPPSYSPPPPTYL 369
 OY 515 APTTPKEPTTPTEEPAPTTKKAAP---NTPKEAPTTKKEPAPTTPKKEAPTTKETA 571
 DB 370 PPPSSPPPSPPPSPPPTYSPPPPPAASPPPLAPPTYSPPPT--YSPPTTYAOP 427
 OY 572 PTPPKGATTTKEAPATTTPKKAPKELAPTTKKEPTTSDKPAPTTPKGAPTTPKEP 631
 DB 428 P-----LPTYSPPPPAASPPPTYSPPPTYSPPPAVAQPPPTYSPPPAVSP 481
 OY 632 APTTPKEPAPTTPKGAPTTTKEAPATTTPKKAPKELAPTTTKGPTTSDKPAPTTPKE 691
 DB 482 PPSPTYSPPPPQVQPLPPTFSPPPRKILPPPHKQF---RPPPTYGGOPSPPTSP 538
 OY 692 TAPTTPKEAPPT--TPKKAAPTTPPTTPPTSEVSPPTTKKEPTTIHKSDESPELSAE 749
 DB 539 PPRQIHSPPPHMQPRPTTPYGGOPSPPTYSAPPROIHSPPPHKQRPRTTYYQCP 598
 OY 750 PPRKALNSPKRPGVPTTKPATKP 775
 DB 599 PSP-----PTTYSPPSPPP 612

RESULT 8
 ID NPH_MOUSE STANDARD; PRT; 1087 AA.
 AC P19246; O61959;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE NEUROFILAMENT TRIPLET H PROTEIN (200 KDA NEUROFILAMENT PROTEIN)
 DE (NEUROFILAMENT HEAVY POLYPEPTIDE) (NF-H).
 GN NEFH OR NEFH.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9121513; PubMed=3220257;
 RA Julien J.-P., Cole F., Beaudet L., Sidky M., Flavell D., Grosveld F.,
 RA Mushynski W.;
 RT "Sequence and structure of the mouse gene coding for the largest
 neurofilament subunit."
 RT Gene 68:307-314(1988).
 RL (2)
 RN SEQUENCE FROM N.A.
 RX MEDLINE=89089138; PubMed=3145094;
 RA Sheldman P.S., Carden M.J., Lees J.F., Lazzarini R.A.;
 RT "The structure of the largest murine neurofilament protein (NF-H) as
 revealed by cDNA and genomic sequences."
 RT Brain Res. 464:217-231(1988).
 RN (3)
 RP SEQUENCE FROM N.A.
 RC STRAIN=SWISS WEBSTER; TISSUE=Brain;
 RA Carden M.J.;
 CC Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
 CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
 CC NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT
 CC SUBSERVED BY THE TWO SMALLER NF PROTEINS.
 CC -1- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P. NEFH IS
 CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
 CC THOUGHT THAT PHOSPHORYLATION OF NEFH RESULTS IN THE FORMATION OF
 CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
 CC OF AXONAL CALIBER.
 CC -1- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
 CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H). THE
 CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
 CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
 CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
 CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN IN POSITIONS 534

RX MEDLINE-99018118; PubMed-979793;
 RA Glockner G., Scherer S., Schattevoy R., Boright A., Weber J.,
 RA Teul L.C., Rosenthal A.;
 RT "Large-scale sequencing of two regions in human chromosome 7q22:
 RT analysis of 650 kb of genomic sequence around the EPO and CMT1 loci
 RT reveals 17 genes.";
 RL Genome Res. 8:1060-1073(1998).
 RN [2]
 RP SEQUENCE OF 2338-2700 FROM N.A.
 RC TISSUE-Testis;
 RX MEDLINE-97271566; PubMed-9126492;
 RA Gao Z., Hartum T., Garbers D.L.;
 RT "Chromosome localization of the mouse zonadhesin gene and the human
 RT zonadhesin gene (ZAN).";
 RL Genomics 41:119-122(1997).
 CC -1 FUNCTION: BINDS IN A SPECIES-SPECIFIC MANNER TO THE ZONA PELLUCIDA
 CC OF THE EGG. MAY BE INVOLVED IN GAMETE RECOGNITION AND/OR
 CC SIGNALING.
 CC -1 SUBUNIT: PROBABLY FORMS COVALENT OLIGOMERS.
 CC -1 SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN, EXCLUSIVELY ON THE
 CC APICAL REGION OF THE SPERM HEAD (BY SIMILARITY).
 CC -1 TISSUE SPECIFICITY: IN TESTIS; PRIMARILY IN HAPLOID SPERMATIDS.
 CC -1 DOMAIN: THE MAM DOMAINS PROBABLY MEDIATE SPERM ADHESION TO THE
 CC ZONA PELLUCIDA.
 CC -1 DOMAIN: DURING SPERM MIGRATION THROUGH THE REPRODUCTIVE TRACTS,
 CC THE MUCIN-LIKE DOMAIN MIGHT INHIBIT INAPPROPRIATE TRAPPING OF
 CC SPERMATOZOA OR PROMOTING ADHESION TO THE OVIDUCTAL ISTHMUS.
 CC -1 DOMAIN: THE VWFD DOMAINS 2 AND 3 MAY MEDIATE COVALENT
 CC OLIGOMERIZATION (BY SIMILARITY TO HUMAN INTESTINAL MUCIN MUC2).
 CC -1 SIMILARITY: CONTAINS 3.5 MAM DOMAINS.
 CC -1 SIMILARITY: CONTAINS 4.5 VWFD DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF053356; AAC78790.1; -
 DR EMBL: U83191; AAC51208.1; -
 DR MIM: 602372; -
 DR InterPro: IPR000561; EGF-1like.
 DR InterPro: IPR000998; MAM.
 DR InterPro: IPR002965; P-rich_extensn.
 DR InterPro: IPR002919; TIL.
 DR InterPro: IPR003328; TILA.
 DR InterPro: IPR001007; VWFC.
 DR InterPro: IPR001846; Vwd.
 DR Pfam: PF00629; MAM; 3.
 DR Pfam: PF01826; TIL; 4.
 DR Pfam: PF02345; TILA; 4.
 DR Pfam: PF00094; vwd; 4.
 DR PRINTS: PR01217; PRICHEXTENS.
 DR SMART: SM00137; MAM; 2.
 DR SMART: SM00214; VWC; 1.
 DR SMART: SM00011; VWC_def; 3.
 DR SMART: SM00216; VWD; 3.
 DR PROSITE: PS01186; EGF_2; 3.
 DR PROSITE: PS00740; MAM_1; 1.
 DR PROSITE: PS00060; MAM_2; 4.
 KW Glycoprotein; Transmembrane; Cell adhesion; Repeat.
 FT NON_TER 1
 FT DOMAIN <1 109 MAM 1.
 FT DOMAIN 112 136 MAM 2 (PARTIAL).
 FT DOMAIN 161 326 MAM 3.
 FT DOMAIN 322 446 MAM 4.
 FT DOMAIN 483 951 66 X HEPTAPEPTIDE REPEATS (APPROXIMATE)
 FT DOMAIN 953 1065 (MUCIN-LIKE DOMAIN).
 FT DOMAIN 1066 1454 VWFD 1 (PARTIAL).
 FT DOMAIN 1455 1861 VWFD 2.
 FT DOMAIN 1455 1861 VWFD 3.

FT DOMAIN 1862 2292 VWFD 4.
 FT DOMAIN 2293 2684 VWFD 5.
 FT DOMAIN ? EGF-LIKE.
 FT CARBOHYD 74 74 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 403 403 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1023 1023 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1099 1099 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1618 1618 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1737 1737 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1832 1832 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1878 1878 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2136 2136 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2505 2505 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2374 2379 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT NON_TER 2700 NNOKMA -> RAGPGR (IN REF. 1).
 FT SEQUENCE 2700 AA; 293013 MW; 80B60CC0B12277B1 CRC64;

Query Match 8.2%; Score 555.5; DA 1; Length 2700;
 Best Local Similarity 32.6%; Pred. No. 7e-16;
 Matches 199; Conservative 64; Mismatches 240; Indels 107; Gaps 32;

QY 123 PINRPSPILPNSDTSKSLTVNKKETVETKETTNNKOTSDGKKTSAKETOSIEKT 182
 DB 445 PVKVLPELPVSVSS-----TGSETTGLTENPTISTK-----KPTVSIKRP 487
 QY 183 SAKDLAPTSVLAKPPPKAETTTKGPALTTPKEPTTPPEKPASTTPKPEPT-----PTTI 237
 DB 488 SVTTEKPT-----VPKEKPTPEKPTISTEK-----PTISEKPNMSEKPTIPSEKPTIL 539
 QY 238 KSAPTPKPEAPTTTKSAPTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTT 296
 DB 540 TEKPTTPE-KPTIPSEKPTISTEKPTVPEE--PTTPEETTTVEEVPITPEKPSIPT 596
 QY 297 TPKEAPTTPKKPAPTTPKEPAATTPKEPTTPKEPTTPKEPAATTPKEPAATTPKEPAATTP 356
 DB 597 --EKPSIPTEK--PTISMEETIISTEKPTICEKPTITEK--PTIPEKSTISPEK- 647
 QY 357 APTTPKEAPTTTPKEPAATTTKEPSPTTPKEPAATTTKSAPTTKEBPATTTKSAPTTPK 416
 DB 648 -PTTPE-KPTIPEKPTISTEKPTIPEK-PTISEKPTIPEKTIPT--EKPTIPT 701
 QY 417 EPSPTTKEBPATTPKEPAATTPKKPA-----PTPKKEBPATTPKEPAATTTKKAPAP 471
 DB 702 EKPTISTEE-PTTPEETIISTEKPSIPMEKPTIPEETTTVEETIISTEKTIPT-- 757
 QY 472 KEBPATTPKEPTAPT-----TPKKIPTTPKEKLA-----PTTPEKPAATTPPELAATTP 521
 DB 758 EKPTISTEKPTIPEKPTISTEKTIPT--EKLIPTKEPTIPIEETIISTEKL--TIPT 814
 QY 522 PPTTPEEPAPTTPKAAAPNTPKEPAATTPKEPAATTTKEBPATTTKEBPATTTKEBPATTT 581
 DB 815 KPTIPEKPTISTEK-----PTIPE-KPTIPE-----ETTTSTEKLTIPT 855
 QY 582 TKEBPATTPKKAPKELAPTTKEPTST-----SDKPAATTPKGAATTPKEPAATTT 635
 DB 856 --EKPTIPEKTIPTKEPTISTEKPTIPEKTIPTKEKTIPTKEKTIPTKEKTIPT--EKLALR 911
 QY 636 PKEPATTPKGAATTTLEKPAATTPPKKAPKELAPTTTGGTSTTSKPAATTPKEK--A 693
 DB 912 PPHPSPTA-TGLALVMSPHAPSTPMSTV---ILGTTTSSSTGMSGP-PNARYESCAC 966
 QY 694 PTPKEPAAT 703
 DB 967 PASCSPRPS 976
 RESULT 7
 ID EXTN_TOBAC STANDARD; PRT; 620 AA.
 AC P13983;
 DT 01-JAN-1990 (Ref. 13, Created)
 DT 01-JAN-1990 (Ref. 13, Last sequence update)

OY 922 -----TSMPELNPTSRFAEAMLOTTRNPTNSKLIVENPKSEDAAGAGETPHM- 973
 DB 1454 YLNGYDPCGSRPRERNITRAAAVIF-----AKLL-----GADESGAASASPSYSD 1498
 OY 974 LMRP-----VEMPE-----VTPDMADLPVPVNOGI----- 999
 DB 1499 LADTHMAAAAIKFRATSGLEKGYRPGDGTNITRAEFATVVLHFLTVKQGEIMSKIA 1558
 OY 1000 ---LIPMLSDETNINCNG-----KPVGLTTL-----RNGTLVAFRGHYFMMLSPSPSPS 1046
 DB 1559 TIDISNPKRDP-----CVGHMAOEFLKTLTSLGYISGYPGT-----FKPON 1600
 OY 1047 PARRTIEWGIPSPIDTFTTRCNCEGKTEFFKD--SQWRFTNDIKD 1091
 DB 1601 YIKRSESV-----ALINRALERGPLNCAKPLFPDVNESYNAF-GDIND 1642

RESULT 3
 AMYL_YEAST STANDARD: PRT: 1367 AA.
 ID AMYL_YEAST
 AC P08640; P08068;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE GLUCOMYLASE S1/S2 PRECURSOR (EC 3.2.1.3) (GLUCAN 1,4-ALPHA-
 GN GLUCOSIDASE) (1,4-ALPHA-D-GLUCAN GLUCOHYDROLASE).
 STAI OR STAI2 OR MAL5 OR YIR019C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_Taxid=4932;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=5288C / AB972;
 RA Barrett B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
 RA Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
 RA Guelles S., Hamlyn N., Horsnell T.S., Hunt S., Jagsels K., Jones M.,
 RA Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
 RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
 RA Walsh S.V., Whitehead S.;
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 1-242 AND 762-1331 FROM N.A.
 RX MEDLINE=87194600; PubMed=3106330.
 RA Yamashita I., Nakamura M., Fukui S.;
 RT "Gene fusion is a possible mechanism underlying the evolution of
 RT STAI.",
 RL J. Bacteriol. 169:2142-2149(1987).
 RN [3]
 RP SEQUENCE OF 1-31 FROM N.A.
 RC STRAIN=SPX101-1C;
 RX MEDLINE=89031230; PubMed=3141213;
 RA Pardo J.M., Ianez E., Zalacain M., Claros M.G., Jimenez A.;
 RT "Similar short elements in the 5' regions of the STAI2 and SGA genes
 RT from Saccharomyces cerevisiae.",
 RL FEBS Lett. 239:179-184(1988).
 CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL 1,4-LINKED ALPHA-D-
 CC GLUCOSE RESIDUES SUCCESSIVELY FROM NON-REDUCING ENDS OF THE CHAINS
 CC WITH RELEASE OF BETA-D-GLUCOSE.
 CC -1- SIMILARITY: TO S.POMBE SPBC215.13.
 CC -1- SIMILARITY: TO S.POMBE SPCC285.13C.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb.ch).
 CC EMBL: 238061; CAA86176.1; -
 DR EMBL: M16164; AAA35014.1; -

DR EMBL: M16165; AAA35015.1; -
 DR EMBL: X13857; CAA32069.1; -
 DR PIR: B26877; B26877.
 DR PIR: A26877; A26877.
 DR PIR: S48478; S48478.
 DR SGD: S0001458; MOC1.
 KW Hydrolyase; Glycosidase; Polysaccharide degradation; Glycoprotein;
 KW Signal; Multigene family.
 FT SIGNAL 1 21
 FT CHAIN 22 1367
 FT DOMAIN 210 1367
 FT CARBOHD 817 817
 FT CARBOHD 874 874
 FT SEQUENCE 1367 AA; 136110 MW; 91C00E2DBD61AA9D CnC64;

Query Match 11.7%; Score 798.5; DB 1; Length 1367;
 Best Local Similarity 27.9%; Pred. No. 5.9e-26;
 Matches 313; Conservative 108; Mismatches 523; Indels 177; Gaps 50;

OY 44 DDAQCKKRYDK--CCPDYESFCAEVK-DNKKNRKKKPPRPVYDE--AGSGLDNGDF 97
 DB 174 DLSTGCNNYDNOGHSGQDFPGFYWNIDCNDNCGGKTSSTTSSESTTSSTTSST 223
 OY 98 KVTTPDSTQHNKVSSTPKITAKPLNP-----RSLRPNSDTSKETSILVKKET 149
 DB 234 TSSSTSSSTTSSTTSSESTTSSTTSSTTAAPPTTSCTEKPRPTTSCTEKPRPHND 293
 OY 150 VETKETTITTKQNTSDKEKTSKETSIEKTSANOLAPTSVLAKPRKAEY--TKG 207
 DB 294 PCFKKKTTSK-TCY--KTTTPVPTPS--SSTTESSAPV-----PTSSSTSSSA 342
 OY 208 PALITPKE-----PTPTTKEPASTTPKEPTPTTKSAP--TTPKEAPPTTKSAP 257
 DB 343 PVTSSTTESSAPVPTPSSTTESSAPVTSSTTESSAPVTSSTTESSAPVPTPSST 402
 OY 258 TPKEAPPTTKKEPAPPTTKKEPAPPTTKKAPPTTKKAPPTTKKAPPTTKKEP 317
 DB 403 TESSAPVTS-----TTESSAPVTS--STTESSAPVTS--SSTTESSAPVTSST 451
 OY 318 APPTPKKEPP--TTPKEPAPPTTKKEPAPPTTKKAPPTTKKAPPTTKKEPAP--TTPK 370
 DB 452 ESSAPVPTPSSTTESSAPVTS--SSTTESSAP--VTPSSSTTESSAPVTSSTTS 507
 OY 371 EPAPPTTKKEPPTTKKEPAPPTTKKAPPTTKKEPAP--TTPKAPPTTKKEPPTTK 426
 DB 508 SSAPVPTPS--SSTTESSAPAPPTPSSTTESSAPVTSSTTESSAPVPTPSSTTS 566
 OY 427 AP--TTPKEPAP--TPKPAPTTKKEPAPPTTKKEPAPPTTKKAPPTTKKEPAPPTTKE 481
 DB 567 TPVTSSTTESSAPVPTPSSTTESSAPVPT--PSSSTTESSAPAPPTPSSTTS 622
 OY 482 TAPPTTKKITPTTPEKLAPTTPEKAPPTPEELAPT--TPEPTPTTPEEAP--TTP 535
 DB 623 SAPVT--SSTTESSAP--VTPSSSTTESSAPVPTPSSTTESSAPVPTPSSTTE 677
 OY 536 KAAAPVTPKEPAPPTTKKEPAPPTTKKEPAPPTTKKAP--TPKGAPPTLKKEPAPPTTKP 594
 DB 678 SSAPVT--SSTTESSAPVT--SSTTESSAPVT--SSTTESSAPVT--SSTTESSAPVT 729
 OY 595 AKKEAPPTTKKPTSTSDKAPPTPKGTAPPTKEPAP--TTPKEPAP--TPKGAPPT 650
 DB 730 TESSAPVPTPS--SSTTESSAPVTSSTTESSAPVPTPSSTTESSAPVPTPSSTTE 788
 OY 651 TLKEPAPPTPKKPAKKEAPPTTKGPTSTSDKAPPTTKKAPPTTKKEPAP--TTP 705
 DB 789 SSAPVPTPSSTTESSAPVPTPSSTTSITSSAPSTPSSSTTESSAPVPTPSSTTE 848
 OY 706 KKPAPPTPEPTPTTSEVSTPT--TTPKEPTIHKSPDSTPELSKEPPLKLESPK 760
 DB 849 SSAPVSSSTTESSAPVPTPSSTTSITSSAPSI--PSSSTTESSTGT--VTTPSSSK 904
 OY 761 EPVPTTTPAATPKPEMTTKKDKTTERDLATTPETTTAAPKMKETATTEKTESKIT 820


```

OY 565 T-TPKETAPTPKGAAPTLKEBAPTPPKKPAKELAPTTTKPESTSDKAP--TPKKG 622
DB 1958 TPTGTOTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 2017
OY 623 TAPTTKEBAPTPPKBAPTPPKGAAPTLKEBAPTPPKKPAKELAPTTTKGPTSTSD 682
DB 2018 TTTTTPPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 2076
OY 683 KPAAP-TTPKETAPTPPKBAPTPPKKPAAPTPPTPTPTPTPTPTPTPTPTPTPT 741
DB 2077 TQPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 2132
OY 742 STEPLSAPPKALNSPKAP-----GVPT-TKTPAATKPEMTTAKDKTERDL 790
DB 2133 TTTTTPPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 2192
OY 791 R---TPP--ETTPAAPKMT--KETATTTETKESKITAATTOVSTTODTTPFKITL 842
DB 2193 QTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 2248
OY 843 KTTTLAKRVTTT-KKITTTEIMKKBEETAKPKDRATNSKATPKPKAP-KKPTST 900
DB 2249 -TTTTPPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 2307
OY 901 KKPMPKPVKPKPTPTPTPKMTSTMPKELNP--TSRIAEMLQTTTR-PMOTPSKILEV 956
DB 2308 QTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 2367
OY 957 NPKSEDAGAGETPHMLLRPHVMEPYTP 986
DB 2368 TPTPTPTG---TQTP--TTPTTTTTTPTVTP 2392

RESULT 2
SLIP1_CLOTM STANDARD; PRT: 1664 AA.
ID SLIP1_CLOTM
AC 006852;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE CELT SURFACE GLYCOPROTEIN 1 PRECURSOR (OUTER LAYER PROTEIN B) (S-LAYER
DE PROTEIN 1).
DE OLBP.
OS Clostridium thermocellum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1515;
RN SEQUENCE FROM N.A.
RC STRAIN=NCIB 10682;
RX MEDLINE=93209931; PubMed=8458832;
RA Fujino T., Beguin P., Aubert J.-P.;
RT "Organization of a Clostridium thermocellum gene cluster encoding the
RT cellulosomal scaffolding protein ClpA and a protein possibly involved
RT in attachment of the cellulosome to the cell surface.";
RT J. Bacteriol. 175:1891-1899(1993).
RL
CC -1 SUBUNIT: ASSEMBLED INTO MONO-LAYERED CRYSTALLINE ARRAYS.
CC -1 SIMILARITY: CONTAINS 4 S-LAYER HOMOMOLOGY (SLH) DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: X67506; CAA47841.1;
CC InterPro: IPR001119; SLH.
CC Pfam: PF00395; SLH. 3.
CC PROSITE: PS01072; SLH_DOMAININ. 2.

```

```

KW Cell wall: S-layer; Signal; Repeat.
FT SIGNAL 1 28
FT CHAIN 29 1664
FT DOMAIN 36 763
FT REPEAT 36 191
FT REPEAT 207 363
FT REPEAT 409 565
FT REPEAT 607 763
FT DOMAIN 771 1377
FT DOMAIN 1378 1449
FT DOMAIN 1453 1494
FT DOMAIN 1495 1565
FT DOMAIN 1566 1625
FT DOMAIN 1626 1646
SO SEQUENCE 1664 AA; 178194 MW; 5F39695BA9FE74B CRC64;

Query Match 13.9%; Score 950; DB 1; Length 1664;
Best Local Similarity 31.0%; Pred. No. 5,1e-32;
Matches 312; Conservative 104; Mismatches 361; Indels 230; Gaps 55;

OY 193 VLAQTPP-KAETTKGAPALTPEKEPTPTPKBAPSTTPKEPTPTTKSAPTPKBPAPTT 251
DB 758 VVIQAPAPKAASDEPIPTDPSDEPTPS-----DEPTPS--DEPTPSDEPTSD 804
OY 252 TKSAPTPKBAPTPPKBAPTPPKBAPTPPKBAPTPPKBAPTPPKBAPTPPKBAPTPPK 307
DB 805 EPPPTSEETPTPTDTPSDDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 864
OY 308 KPAATTPKBPAPTPPKBAPTPPKBAPTPPKBAPTPPKBAPTPPKBAPTPPKBAPTPPK 366
DB 865 SDEPTPSDEPTPS--DEPTPS--DEPTPS--DEPTPS--DEPTPSDEPTPSDEPT 919
OY 367 TTKKEPA--PTTKESAPT--TKKEBAPTTS--APTTPKEAPT--TTKSAPTTPKES 419
DB 920 SDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 979
OY 420 ---PTTKBPAPTPPKBAPTPPKBAPTPPKBAPTPPKBAPTPPKBAPTPPKBAPTPPK 474
DB 980 PSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 1039
OY 475 APT--TPKETAPTPPKKLPPTPEKAPTPPKBAPT--TPEELAPTPPEPTPT-TP 527
DB 1040 TPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 1094
OY 528 EEPAPTPPKAAPTTPKBPAPTPPKBAPTPPKBAPTPPKBAPTPPKBAPTPPKBAPTPPK 584
DB 1095 EEPPTDTPSDDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSD 1154
OY 585 EPAPTPPKBAPKELAPTTKEPT-STTSKBPATTPKGAAPTPPKBAPTPPKBAPTPPK 643
DB 1155 EPTPS--DEPTPS--EPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPS--DEPTPS- 1208
OY 644 PKGAAPTLKBPAPT-TPKBPAPKELAPTTTKGPTSTSDKAPAT--TKETAAPTTPKE 699
DB 1209 ---DEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSD-EP 1253
OY 700 PAPTPPKBPAPT-TPBPPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 758
DB 1254 PTPS--DEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPS--SDEPTPSDEPT 1305
OY 759 PKEBPVPTTKPAPAT--KPEMTTAKDKTERDLATPEETTAAPKMTKEATTTTEKTE 816
DB 1306 SDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSD---EPTPSDEPTPSDEPTPSDEPTPS 1361
OY 817 SKITATTTQVSTTTQDTPTPEKITTLLKTTTLAKVTTTKKTTTTEIMKKBEETAKPKDR 876
DB 1362 DEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 1396
OY 877 ATNSKAT-----TPKQKPTKAP-KKPTSTKPKMTKPKPKMTPTPKMK----- 921
DB 1397 SGGSGSGGGGGGGGGGGGGTPTPTPTSTKPTSTPAP---TEIEPTPSDVGAIGEHRA 1453

```

FT	CARBOHYD	4339	4339	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	4351	4351	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	4362	4362	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	4373	4373	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	4422	4422	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	4438	4438	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	4502	4502	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	4616	4616	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	4627	4627	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	4752	4752	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	4787	4787	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	4881	4881	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	4888	4888	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	4955	4955	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	4970	4970	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	5019	5019	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	5038	5038	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	5069	5069	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CONFLICT	1351	1351	H -> I (IN REF. 3)		
FT	CONFLICT	1412	1412	T -> S (IN REF. 3)		
FT	CONFLICT	1449	1449	L -> P (IN REF. 3)		
FT	CONFLICT	1504	1504	M -> T (IN REF. 3)		
FT	CONFLICT	1504	1504	G -> S (IN REF. 2)		
FT	CONFLICT	4192	4192			
SQ	SEQUENCE	5179	AA: 540295	85CD7511EB9A5663	CRC64:	

	Query Match	17.28;	Score 1171;	DB 1;	Length 5179;	
	Best Local Similarity	32.88;	Pred. No. 1.Se-40;			
	Matches 364;	Conservative 57;	Mismatches 461;	Indels 228;	Gaps 41;	
QY	38 EGRGECDDAAC-----CKKYDK-----CCPDYE--SFCACVKNKKNRKTKRPKPVPV	84				
Db	1350 OHGQVQCDDVASVGFICKNEDQFGNGPRLCYDXYIRNVCCPMH-----KCITTPSPPT	1403				
QY	85 VDEAGSGLDNDDFKVTTPDPDISTOHNKVYSIPKITTAKPINRPSLPPNSDTSKS--TSL	142				
Db	1404 TTP-----SPPPTTTT-----TLPPITTSPPTTTTTTPPTTTPSPPTTT	1445				
QY	143 TVNKETVEERK--ETTUTNNKOTSDIGKEKTSAKETOSIEKTSAKDLAPISKVLAETPK	200				
Db	1446 TTPLPTTTPSPISTTTTTPPTTTPSPPTTTPSPPTT-----PSP	1487				
QY	201 AETTTKGPAITPKPEP--TPTPKEPASTPKEDPTTIKSAPTTPKEPATTTTKSAPT	258				
Db	1488 TTTTTPPTTPPTTPBPMTPTTP--PASITTLPTP--TTPSBPTTTTTTPPTTTPSPPT	1544				
QY	259 PKEPAPTTPKEPATTTPKEPAPTTPKBPATTTKSAPTTPKEPATTPKBPATTPKEPA	318				
Db	1545 TPITPPTSTTLTPPTTTPSPPTTTTTTPPTTTPSPPT-----TTPSPPTTTTTPPP	1598				
QY	319 PTPPKERPPTTPKEPATTPKEPATTPKEP-----APTAKKBPATTPKEPATTPKE	371				
Db	1599 TTPTPSPPTTTTTTPPTTTPSPPTTTPPTTPPTSTTLTPPTTTPSPPTTTTTTPPTTTPS	1658				
QY	372 PAPPTKEPSTTPPKEPAPTTPKSAPTTPKEPATTTTKSAP---TTPKESPTTTKBPAP	428				
Db	1659 PPTTTTSSPTTTTTTPPTTTPSPPTTTPSPPTTTTTTPPTTTPSPPTTSSPI--TTTTTPSS	1717				
QY	429 TTPKEPATTPPKBPATTPPKEPAPTTPKEPATTPPKBPATTPKBPATTPKEPATTPP--	KETA	483			
Db	1718 TTPSPSPPTTMTPSPPTTTPSPPTTWTLLPPTTSSPLTTTLPLPISITTPPTSPTT	1777				
QY	484 PTTP-----	KLTP-----	492			
Db	1778 PTTPCVPCLMNTGMLDSGKPNFKPGDTELIGDYCGPGMAANISCRATWYPDPVLIGOLG	1837				
QY	493-----	TPEKLAPTTPKEPATTP	510			
Db	1838 QTVVCADVGLICKNEQKRGCVIIPAFCALNEYENNOCCCACTVQTPTMTTTTTTENPRPT	1897				
QY	511 PEELAPTTPPEPPTP--TPEEPAPTTPKAAPNPKEPATTPPKP-----APTTPKEPA	564				
Db	1898 TTPTTTTTTPPTTPPGTQPTTTTTPTTTTTTTTTTTTTPPTTPPTTPGTQPTTTTTPPTTTP	1957				

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:32:21 ; Search time 62.75 Seconds
(without alignments)
742.061 Million cell updates/sec

Title: AA3
Perfect score: 6814
Sequence: 1 MAKKTLPIVLLLLSVFVIO.....AAAITRSGQTLSKWNVNC 1270

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues
Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Swissprot_39:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

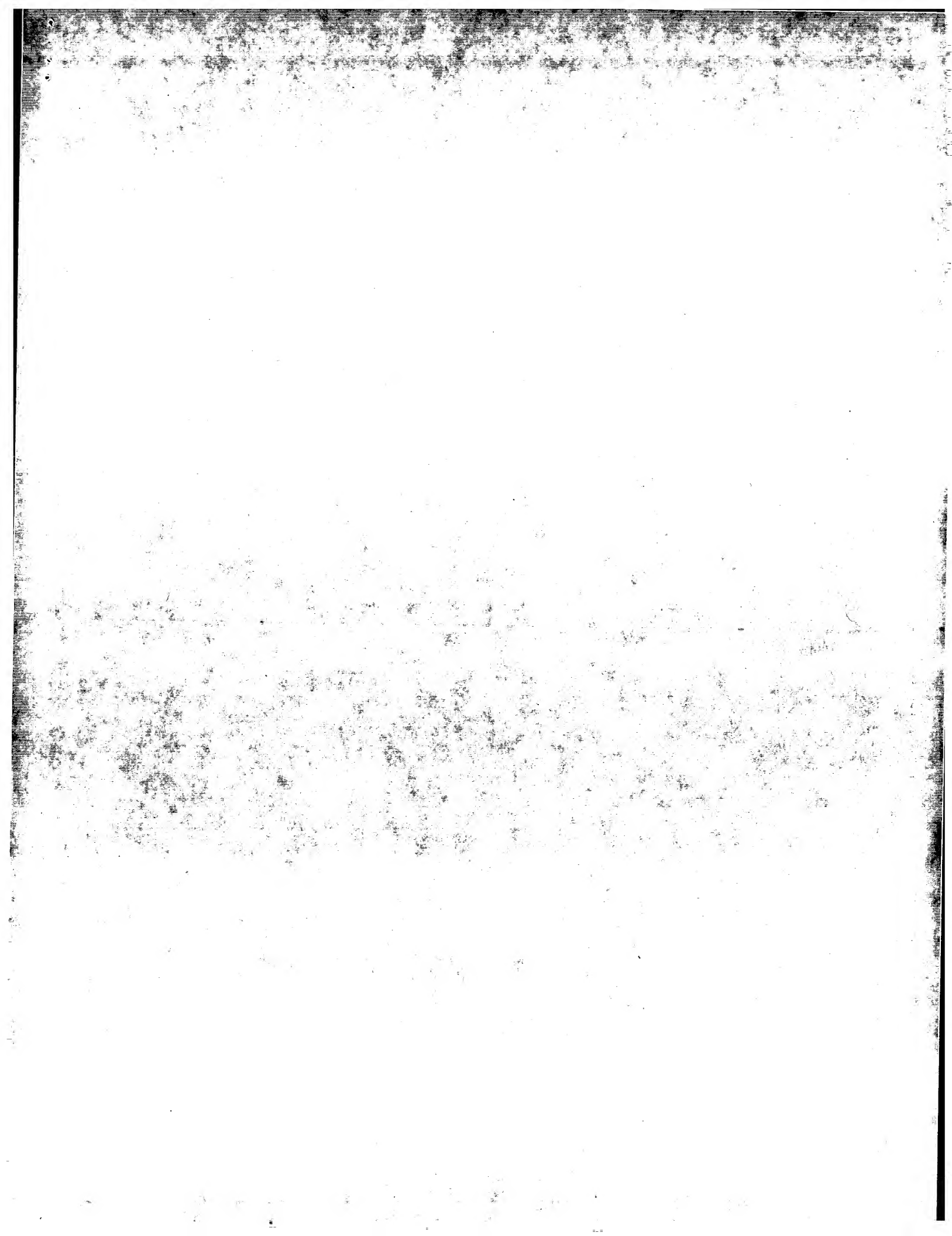
SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	1171	17.2	5179	1 MUC2_HUMAN	Q02817 homo sapien
2	950	13.9	1664	1 SLPI_CLOTH	Q06852 clostridium
3	798.5	11.7	1367	1 AMYL_YEAST	P08640 saccharomyc
4	651	9.6	1255	1 MUC1_HUMAN	P15941 h mucin 1 p
5	556.5	8.2	875	1 PPL_MYTED	Q25460 mytilus edu
6	555.5	8.2	2700	1 ZAN_HUMAN	Q9Y493 homo sapien
7	551	8.1	620	1 EXTN_TOBAC	P13983 nicotiana t
8	533	7.8	1087	1 TCNA_TRYCR	P19246 mus musculu
9	530.5	7.8	1162	1 CPN_DROME	P23253 trypanosoma
10	530	7.8	865	1 PPL_MYTCO	Q02910 drosophila
11	518.5	7.6	872	1 MUC1_XENLA	Q25434 mytilus cor
12	509.5	7.5	662	1 RPB1_HUMAN	Q05049 xenopus lae
13	503.5	7.4	1970	1 RPB1_MOUSE	P24928 homo sapien
14	498.5	7.3	1970	1 RPB1_MOUSE	P08775 mus musculu
15	497	7.3	831	1 NFN_RAT	P16844 rattus norv
16	493.5	7.2	467	1 RPB1_CRIGR	P11414 cricetus
17	488.5	7.2	826	1 SSP2_PLAYO	Q01443 plasmodium
18	488	7.2	1020	1 NFN_HUMAN	P12036 homo sapien
19	475.5	7.0	267	1 NFN_MAIZE	P14918 zea mays (m
20	472	6.9	2142	1 BAT2_HUMAN	P48634 homo sapien
21	471.5	6.9	5376	1 ZAN_MOUSE	O88799 mus musculu
22	468.5	6.9	634	1 HWP1_CANAL	P46593 candida alb
23	467	6.9	817	1 VRP1_YEAST	P37370 saccharomyc
24	455	6.7	1161	1 YJ9P_YEAST	P47179 saccharomyc
25	454	6.7	797	1 VGLX_HSVB	P28968 equine heip
26	448.5	6.6	670	1 VGLX_HSVB	Q00130 ictalurid h
27	442.5	6.5	751	1 FPL_MYTGA	Q27409 mytilus gal
28	439.5	6.4	1083	1 T2D3_HUMAN	O00268 homo sapien
29	432.5	6.3	439	1 XPD_XENLA	P17437 xenopus lae
30	432	6.3	3164	1 TEGU_HSV1	P10200 herpes simp
31	430.5	6.3	3421	1 TEGU_HSVB	P28955 equine heip
32	426	6.3	2715	1 TRX2_HUMAN	O9um66 homo sapien
33	424.5	6.2	1125	1 MAP4_MOUSE	P27546 mus musculu

34	424	6.2	307	1	SGS3_DROME	P02840 drosophila
35	424	6.2	1251	1	Y03_CAEEL	O09550 caenorhabdi
36	422.5	6.2	1794	1	YAV1_SCHPO	O10172 schizosach
37	422	6.2	2774	1	MAPA_RAT	P34926 rattus norv
38	421	6.2	1229	1	N121_HUMAN	Q9Y2n3 homo sapien
39	420	6.2	1411	1	TCOF_HUMAN	O13428 homo sapien
40	419.5	6.2	2476	1	ZAN_PIG	Q28983 sus scrofa
41	419	6.1	1185	1	DRPL_HUMAN	P54259 homo sapien
42	411.5	6.0	907	1	DRPL_DBY	P03200 epstein-bar
43	410	6.0	1183	1	DRPL_RAT	P54258 rattus norv
44	410	6.0	2517	1	NCR2_HUMAN	O9y618 h nuclear r
45	407.5	6.0	3256	1	K167_HUMAN	P46013 homo sapien

ALIGNMENTS

RESULT 1	MUC2_HUMAN	STANDARD:	PRT: 5179 AA.
ID	MUC2_HUMAN	Q02817; 014878;	
AC	01-JUN-1994 (Rel. 29, Created)		
DT	01-NOV-1997 (Rel. 35, Last sequence update)		
DT	20-AUG-2001 (Rel. 40, Last annotation update)		
DE	MUCIN 2 PRECURSOR (INTESTINAL MUCIN 2).		
GN	MUC2 OR SMC.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OX	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
NCBI_Taxid=9606;			
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Intestine;		
RX	MEDLINE=94132002; PubMed=8300571;		
RA	Gum J.R. Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;		
RT	"Molecular cloning of human intestinal mucin (MUC2) cDNA.		
RT	Identification of the amino terminus and overall sequence similarity		
RT	to prepro-von Willebrand factor.";		
RL	J. Biol. Chem. 269:2440-2446(1994).		
RN	[2]		
RP	SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.		
RC	TISSUE=Colon;		
RX	MEDLINE=93016075; PubMed=1400449;		
RA	Gum J.R. Jr., Hicks J.W., Toribara N.W., Rothe E.-M., Lagace R.E.,		
RT	Kim Y.S.;		
RT	"The human MUC2 intestinal mucin has cysteine-rich subdomains located		
RT	both upstream and downstream of its central repetitive region.";		
RL	J. Biol. Chem. 267:21375-21383(1992).		
RN	[3]		
RP	SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.		
RX	MEDLINE=91358717; PubMed=1885763;		
RA	Toribara N.W., Gum J.R. Jr., Culhane P.J., Lagace R.E., Hicks J.W.,		
RT	Petersen G.M., Kim Y.S.;		
RT	"MUC-2 human small intestinal mucin gene structure. Repeated arrays		
RT	and polymorphism.";		
RL	J. Clin. Invest. 88:1005-1013(1991).		
CC	- FUNCTION: COATS THE EPITHELIA OF THE INTESTINES, AIRWAYS, AND		
CC	OTHER MUCUS MEMBRANE-CONTAINING ORGANS. THOUGHT TO PROVIDE A		
CC	PROTECTIVE, LUBRICATING BARRIER AGAINST PARTICLES AND INFECTIOUS		
CC	AGENTS AT MUCOSAL SURFACES.		
CC	- SUBUNIT: MULTIMERIC.		
CC	- SUBCELLULAR LOCATION: SECRETED.		
CC	- TISSUE SPECIFICITY: COLON, SMALL INTESTINE, COLONIC TUMORS,		
CC	BRONCHUS, CERVIX AND GALL BLADDER.		
CC	- PTM: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR		
CC	INTERCHAIN DISULFIDE BONDS (BY SIMILARITY).		
CC	- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC AND		
CC	VARIABLE AMONG DIFFERENT ALLELES.		
CC	- SIMILARITY: THE N-TERMINAL DOMAIN SHOWS SOME SIMILARITY TO THAT		
CC	OF SILKWORM HEMOCYTIN.		
CC	- SIMILARITY: CONTAINS 2 WFC DOMAINS.		
CC	- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTEINE KNOT-LIKE DOMAIN (CTCK).		



Mon Apr 29 08:35:18 2002

aa3.rpr

Page 13

```

QY      800  APMKTKETATTEKTESKITATTTQVSTTTQDTP      836
          |      |      |      |      |      |
Db      704  PPVKPPPVQVPPPTPTYSPPPIKPPPVQVPPPTTSPDP      740

```

Search completed: April 26, 2002, 16:31:52
Job time: 707 sec